



Euroopa Liit
Euroopa Sotsiaalfond



Eesti tuleviku heaks

E-kursuse "Statistikapaketi SPSS kasutamine" materjalid

Tartu Ülikoolis õpetatava vabaaine SOSS.02.018 " **Statistikapaketi SPSS kasutamine**" juurde

Aine maht 3 EAP

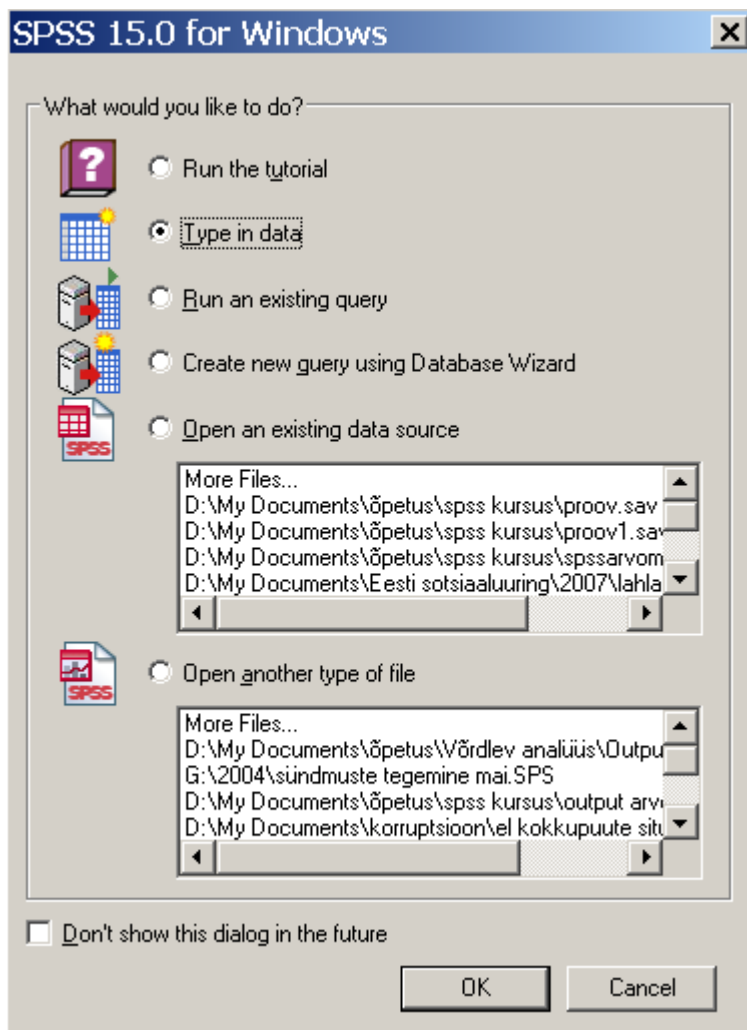
Kadri Rootalu (Tartu Ülikool), 2009

Statistikapaketi SPSS kasutamine

Praktikum 1

Avage SPSS 15.0

Tühja SPSS töölehe avamiseks (näiteks et sisestada oma andmestik) valida „Type in data“ (vt pilt 1).



Juba olemasoleva SPSS faili avamiseks sobib „Open an existing data source“ ning mõne muu SPSS faili avamiseks „Open another type of file“

Valige uue andmestiku sisestamiseks „Type in data“

Pilt 1

SPSS tööaknad

Andmeaken

Variable view ehk tunnuste vaade näitab kõigi andmestikus olevate tunnuste kirjeldusi

Data view ehk andmevaade näitab kõigi andmestikus olevate indiviidide (objektide) tunnuste väärtusi.

Väljundiaken

Väljundiaknas (Output) esitatakse kõik tellitud analüüsitulemused. Uued tulemused paigutatakse eelmiste alla. See aken avaneb siis, kui tulemusi tellida (automaatselt see ei avane).

Programmeerimisaken

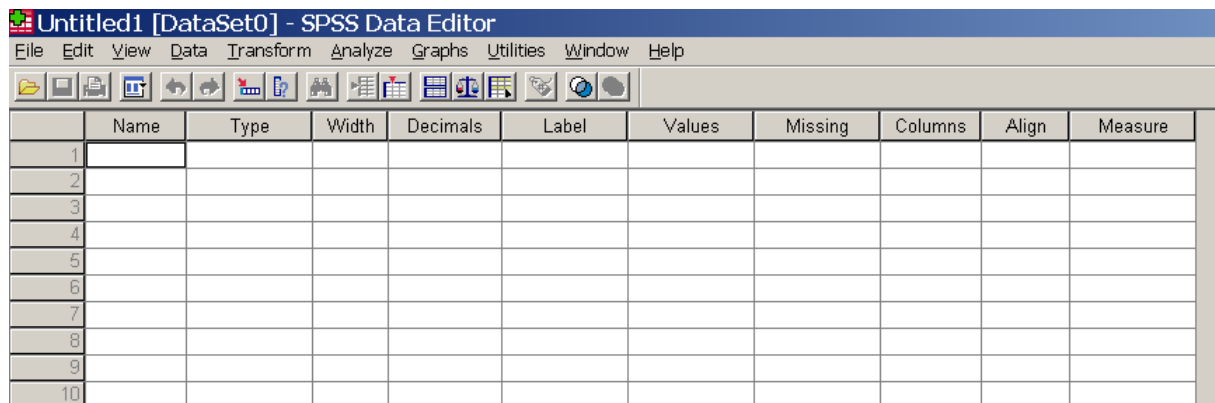
Programmeerimisaknas (Syntax) saab kirja panna käsud analüüside tellimiseks. Akna saab avada, kui soovetakse kasutada programmeerimiskeelt. Automaatselt see ei avane.

Andmestiku sisestamine

NR	SU□U	VANUS	□ARIDUS	RA□VUS	SISS□TUL□K	□UVIALAD
1	mees	30	kesk	eestlane	13500	Sport
2	naine	26	kutse	eestlane	7000	Kirjandus, muusika
3	mees	48	kõrgem	eestlane	□000	□uusika, sport
4	mees	63	kesk	eestlane	3000	Laulmine
5	mees	22	kõrgem	eestlane	15000	Korvpall, muusika
6	naine	3□	kutse	venelane	6000	Kirjandus
7	naine	67	kesk	eestlane	ei ütle	
8	naine	2□	kõrgem	ei oska öelda	10000	Sport, muusika
□	mees	38	kesk	soomlane	11000	Kirjandus
10	naine	43	kõrgem	eestlane	8000	Tennis
11	naine	50	kutse	venelane	4000	Kirjandus, muusika
12	mees	52	kutse	venelane	□500	□ale, kirjandus

Sisestage antud andmestik SPSSi.

Selleks valige andmeaknas „Variable view“, avanev ekraanipilt on esitatud pildil 2.



Pilt 2

□sitatud vaates (tunnuste vaade) märgivad read tunnuseid. Igale tunnusele on vaja omistada:

Name: tunnuse lühinimi (soovitavalt kuni 8 tähemärki, ilma täpitähtede, tühikute ja keerulisemate sümboliteta)

Type: sisestatava informatsiooni tüüp (kas sisestatakse numbreid (Numeric), kuupäevi (Date), teksti (String) vms). Vt ka pilt 3.

Tunnuse tüübi sisestamiseks tuleb:

- 1) Klõkkida vastava tunnuse real ning veerus Type asuval lahtril
- 2) Klõkkida aktiivse lahtri paremas ääres oleval hallil (kolme punktiga) nupul.
- 3) Sisestamise järel vajutada OK

	Name	Type	Width	Decimals	Label	Values
1	nr	Numeric	3	0	Number	None
2	sugu					
3						
4						
5						
6						
7						
8						
9						
10						
11						

Variable Type

☒ Numeric

☐ Comma

☐ Dot

☐ Scientific notation

☐ Date

☐ Dollar

☐ Custom currency

☐ String

Width: 3

Decimal Places: 0

OK

Cancel

Help

Pilt 3

□ idth: numbri□või tähekohtade arv sisestatavas kirjes

Decimals: komakohtade arv numbrikohtade arvu sees

Label: tunnuse nimi, kirjeldus

Values: tunnuse väärtuste koodid ja nende tähendused

Koodide määramise aken avaneb, kui klikkida aktiivse lahtri paremas ääres oleva halli nupukese peal.

Aknas „Value Labels“ (pilt4):

Value: tunnuse kood

Label: koodi tähendus

Iga kood tuleb koos oma tähendusega sisestada ükshaaval. Iga sisestuse järel vajutada „Add“.

Juba sisestatud kirje muutmiseks „Change“ või kustutamiseks „Remove“

Width	Decimals	Label	Values	Missing	Columns	Align	Me
3	0	Number	None	None	8	Right	Scale
2	0	Sugu	None	None	8	Right	Scale

Value Labels

Value Labels:

Value: 2

Label: naine

Add

Change

Remove

1 = "mees"

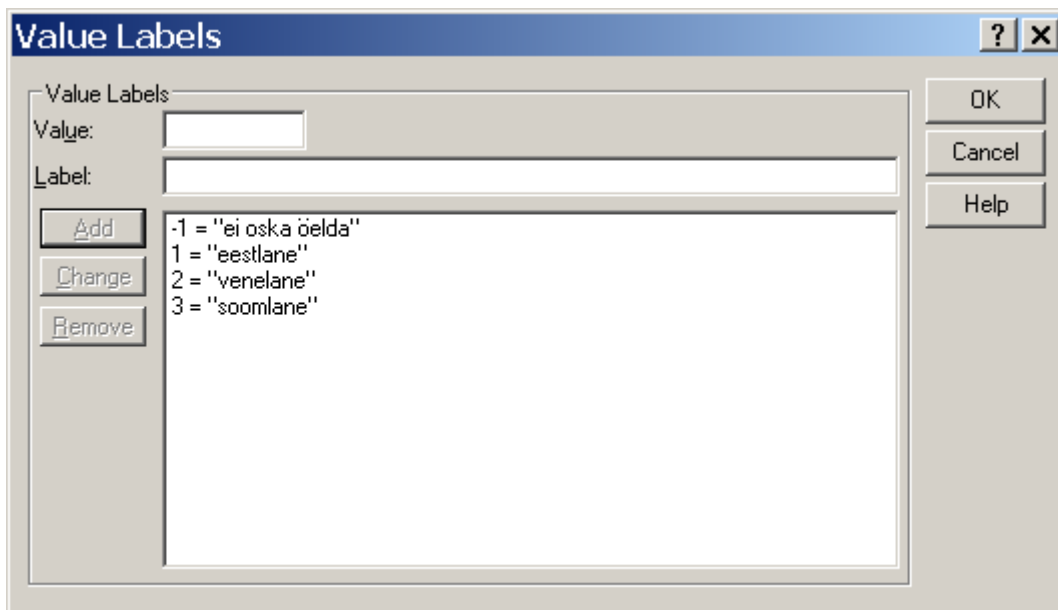
OK

Cancel

Help

Pilt 4.

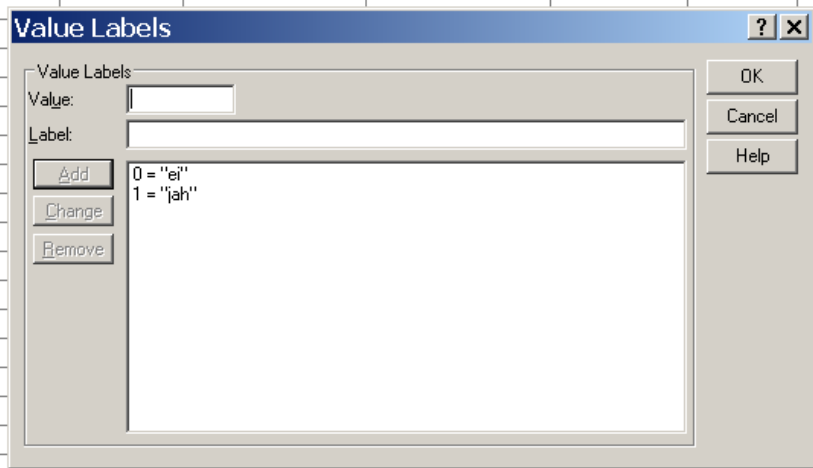
Kui tunnusel leidub puuduvaid väärtusi, siis nende koodid võib samuti sisestada. Näiteks pilt 5 puhul on puuduvaks väärtuseks -1 „ei oska öelda“



Pilt 5.

Tunnused, millele on võimalik anda rohkem kui üks vastus, sisestatakse enamasti iga väärtuse jaoks eraldi koodi teel. Nii võib teha näiteks andmestikus oleva huvialade tunnusega. Sel juhul moodustatakse iga huviala kohta eraldi rida. Näiteks: huvialaks sport (vastusevariantidega jah, kui vastaja on selle märkinud ning ei, kui vastaja pole sporti oma huvialana märkinud). Sellise tunnuse kodeerimist vaata pildilt 6.

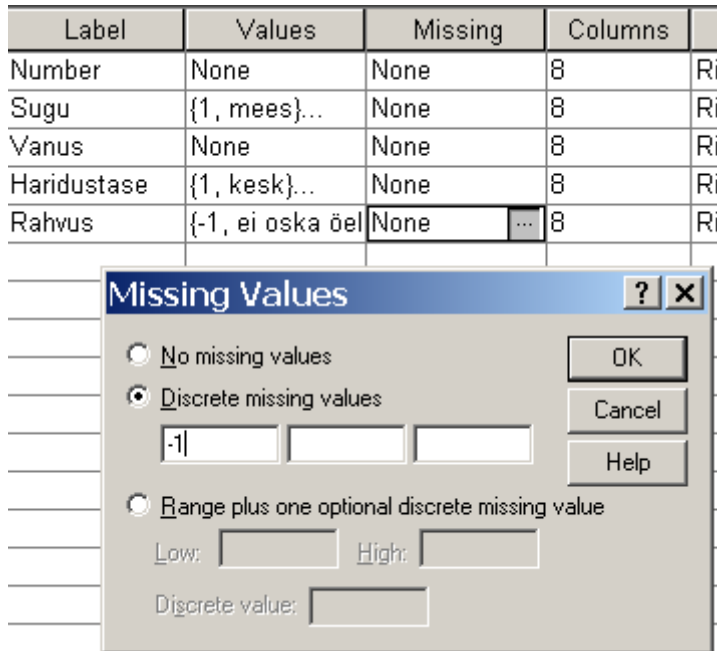
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	
1	nr	Numeric	3	0	Number	None	None	8	Rig
2	sugu	Numeric	2	0	Sugu	{1, mees}...	None	8	Rig
3	vanus	Numeric	3	0	Vanus	None	None	8	Rig
4	haridus	Numeric	3	0	Haridustase	{1, kesk}...	None	8	Rig
5	rahvus	Numeric	2	0	Rahvus	{-1, ei oska öelda}...	-1	8	Rig
6	sisset	Numeric	6	0	Sissetulek (EEK)	{-2, ei ütle}...	-2	8	Rig
7	sport	Numeric	2	0	Huviala: sport	{0, ei}...	None	8	Rig
8	kirjandus	Numeric	2	0	Huviala: kirjandus	{0, ei}...	None	8	Rig
9	muusika	Numeric	2	0	Huviala: muusika	{0, ei}...	None	8	Rig
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									



Pilt 6.

☐ issing: puuduvate väärtuste koodid

Puuduvaid väärtusi on võimalik esitada üksikute väärtustena (Discrete missing values) või siis väärtusvahemiku ja lisaks veel ühe üksiku väärtusena (Range plus one optional discrete missing value). Vt ka pilt 7.



Pilt 7.

Columns: andmeveeru laius

Align: andmeveeru joondus

☐ easure: tunnuse tüüp (nominaaltunnus, ordinaal ☐ ehk järjestustunnus, arvtunnus (Scale))

Sisestatud andmestik võiks välja näha nii, nagu näidatud pildil 8.

*Untitled1 [DataSet0] - SPSS Data Editor										
File Edit View Data Transform Analyze Graphs Utilities Window Help										
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	nr	Numeric	3	0	Number	None	None	8	Right	Scale
2	sugu	Numeric	2	0	Sugu	{1, mees}...	None	8	Right	Nominal
3	vanus	Numeric	3	0	Vanus	None	None	8	Right	Scale
4	haridus	Numeric	3	0	Haridustase	{1, kesk}...	None	8	Right	Ordinal
5	rahvus	Numeric	2	0	Rahvus	{-1, ei oska öelda}...	-1	8	Right	Nominal
6	sisset	Numeric	6	0	Sissetulek (EEK)	{-2, ei ütle}...	-2	8	Right	Scale
7	sport	Numeric	2	0	Huviala: sport	{0, ei}...	None	8	Right	Nominal
8	kirjandus	Numeric	2	0	Huviala: kirjandus	{0, ei}...	None	8	Right	Nominal
9	muusika	Numeric	2	0	Huviala: muusika	{0, ei}...	None	8	Right	Nominal
10										

Pilt 8.

Salvestage saadud andmestik oma eesnimega

☐ ile – Save as

Olemasolevaid andmeid saab ka hiljem muuta ja ümber kodeerida. Samuti saab andmestikku lisada uusi tunnuseid.

- 1) ☐ oodustage sissetulekute baasil uus tunnus, grupeerides sissetulekute vahemikud
1 – alla 5000 krooni
2 – 5000 kuni ☐☐☐☐ krooni
3 – 10000 krooni või rohkem

Transform – Recode - Into Same Variables

Olemasoleva tunnuse ümberkodeerimine selliselt, et vana tunnus asendatakse uuega.

Kuna praegu oleks hea, et ka sissetuleku tunnuse originaal jääks andmestikku alles, siis seda varianti praegu valida ei tohiks.

Transform – Recode - Into Different Variables

Olemasoleva tunnuse ümberkodeerimine selliselt, et andmestikku lisatakse uus tunnus.

Valida ümberkodeerimisele minev tunnus, saata see parempoolsesse aknasse.

Output variable: uue tunnuse nimi

Label: uue tunnuse kirjeldus

Old and New Values: ümberkodeerimise juhise andmiseks (põhineb andmestikus olevatel koodidel)

Old Value: väärtus olemasoleval tunnusel (sellel, mida ümber kodeerite)

New Value: väärtus uuel tunnusel.

Vajutada iga koodi järel “Add”

If: tingimused, mille puhul ümberkodeerimist teostada.

- 3) Tehke sagedustabel uue sissetuleku tunnuse jaotuse kohta.

Analyze – Descriptive Statistics – Frequencies

- 4) Kodeerige ümber rahvuse tunnus (eestlane – mitte-eestlane)

- 5) Tehke uue ja vana tunnuse jaotuste kohta risttabel.

Analyze – Descriptive Statistics – Crosstabs

Row: reatunnuseks vana rahvuse tunnus

Column: veerutunnuseks uus rahvuse tunnus

- 6) Kodeerige ümber vanuse tunnus kolmeks vanusegrupiks enda äranägemise järgi.

- 7) Kodeerige ümber hariduse tunnus kaheväärtuseliseks (kas on kõrgharidus) eestlastel.

Selleks If nupu alt panna tingimus (rahvus ☐ eestlaste kood ☐

Statistikapaketi SPSS kasutamine

Praktikum 2

Andmestike formaadid SPSSis

ri tüüpi faile saab avada
File – Open – Data

Andmestike formaadid

File – sav

File – por

SPSSis saab avada ka teiste analüüsiprogrammide faile. Proovige näiteks avada fail excelist.xls (mitmekesi korraga sama File – xceli faili SPSSis avada ei saa, selleks tuleb eelnevalt faili xcelis avada ja salvestada mingi muu nimega, näiteks excelistomanimi.xls)

Muude SPSS akende formaadid

Väljundakna formaat .spo

Programmeerimisakna formaat .sps

SPSS – n – d

File

Andmestike avamine, ka teistest formaatidest
Andmestike salvestamine

Avage andmestik proov1.sav

1) Tehke sagedustabel vastajate soolise jaotuse kohta
Analyse – Descriptive Statistics – Frequencies

Edit

Edit – Options

Võimaldab muuta SPSS etteantud tehnilisi seadeid (näiteks väljundis programmeerimiskäskude esitamine või keelamine)

Data

Data – Merge files

Saab lisada teisest andmestikust tunnuseid (Variables) või indiviide (Cases).

Data – Split file

Saab jagada andmestiku mitmeks alamandmestikuks mingi tunnuse osas (seejuures muutub indiviidide järjestus andmestikus)

Data – Select cases

Saab valida alamhulga indiviididest

Data – Weight cases

Andmestiku kaalumise

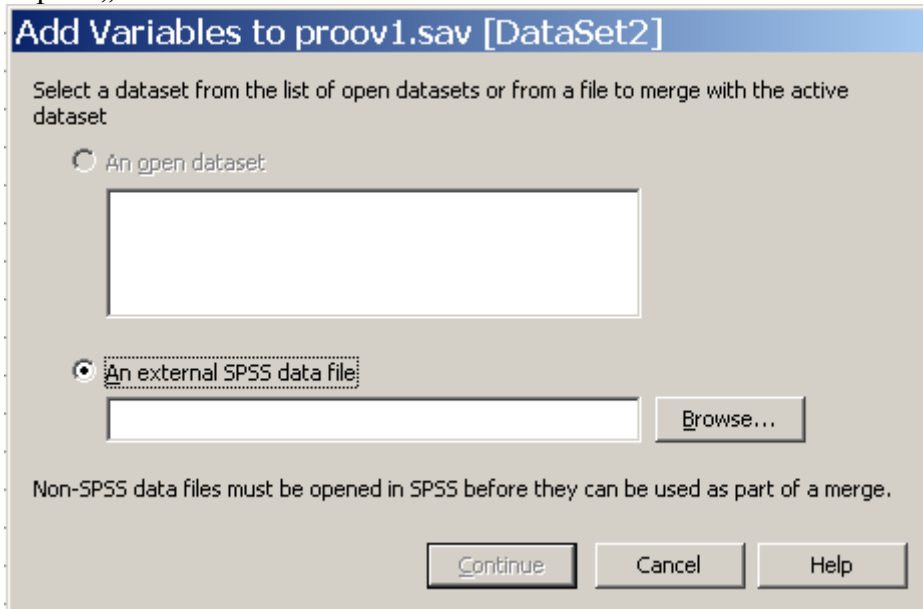
and d

2) Liitke andmestikud proov1.sav, proov2.sav ja proov3.sav.

Data – ☐ erge files – Add Variables

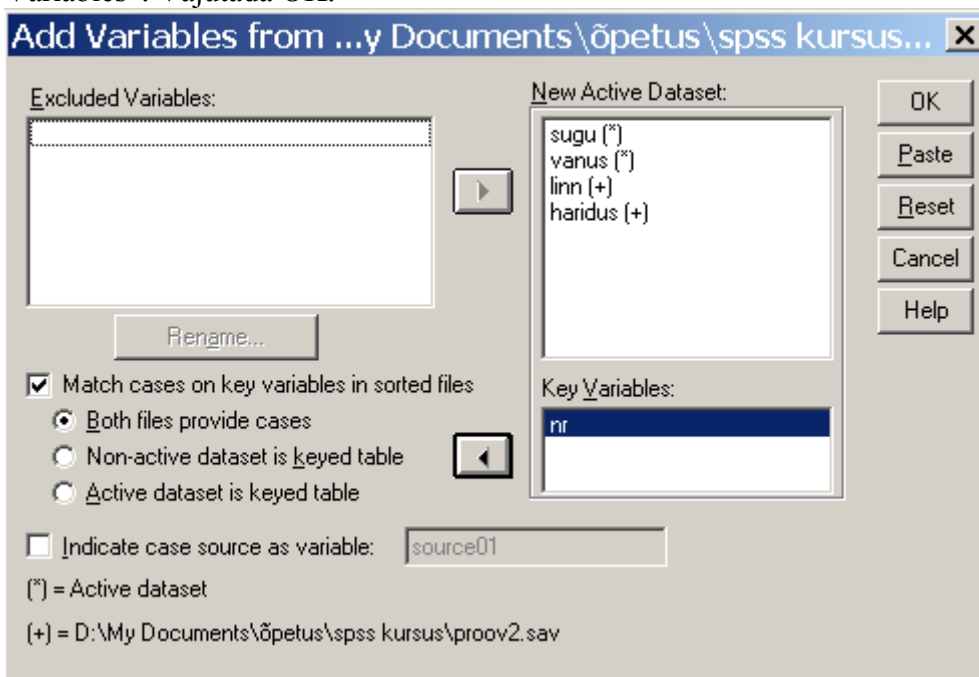
Varasemalt peab olema avatud andmestik proov1.sav.

Seejärel tuleks nupu alt „ browse“ valida andmestik proov2.sav (vt ka pilt 1) ning vajutada nupule „Continue“



Pilt 1.

Avanevas tellimisaknas (pilt 2) on erineva tähistusega (vastavalt plussi ja tärniga) märgitud aktiivse andmestiku (proov1.sav) ning lisatava andmestiku tunnused. Tunnus „nr“ kordub mõlemas andmefailis ja selle tunnuse järgi saab andmefailid ühendada. Selleks märkida ära „atch cases on key variables in sorted files“ ning saata tunnus „nr“ lahtrisse „Key Variables“. Vajutada OK.



Pilt 2.

Sama protseduuri korrates ühendage olemasoleva andmefailiga ka fail proov3.sav.

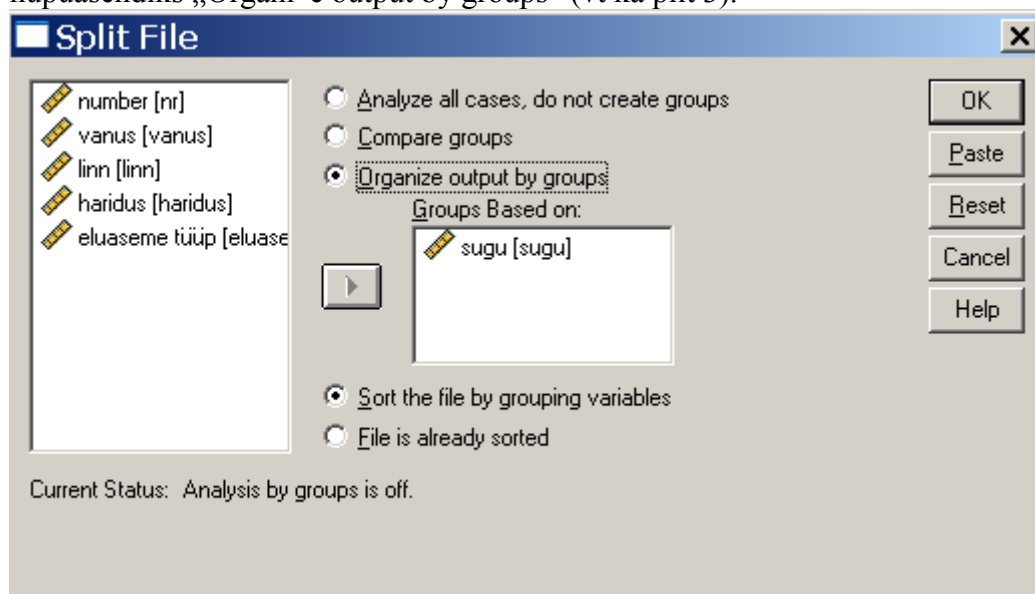
On oluline, et ühendatavad andmefailid oleks sorteeritud ühendatava tunnuse (nr) järgi. Vastasel juhul andmestike ühendamine ei õnnestu.

3) Salvestage saadud ühendandmestik oma eesnimega.

4) Jagage saadud andmestik soo lõikes kaheks.

Data – Split file

Selleks saatke tunnus „sugu“ parempoolsesse lahtrisse „Groups Based on“ ning valige nupuasendiks „Organize output by groups“ (vt ka pilt 3).



Pilt 3

5) Tehke haridustaseme tunnuse kohta sagedustabel (nii et soo lõikes oleksid eraldi tabelid).

Analyse – Descriptive Statistics – Frequencies

6) Tehke haridustaseme tunnuse kohta sagedustabel kõigile vastajatele (selleks võtke ära andmestiku jagamine soo lõikes).

Data – Split file

Soo lõikes jagunemise eemaldamiseks valige „Analyze all cases, do not create groups“

Analyse – Descriptive Statistics – Frequencies

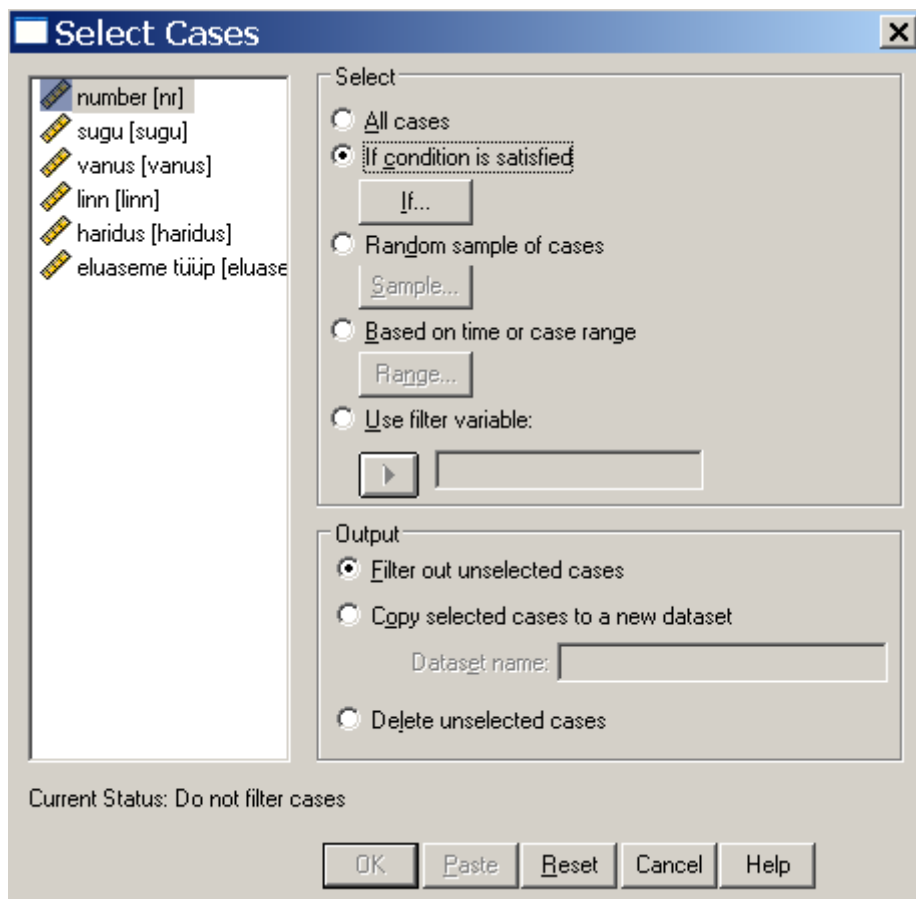
7) Selekteerige filtri abiga ainult Tallinna elanikud.

Data – Select cases

Tellimisaknas valige selekteerimiseks variant „If condition is satisfied“ ning vajutage nupule „If“ (vt. ka pilt 4).

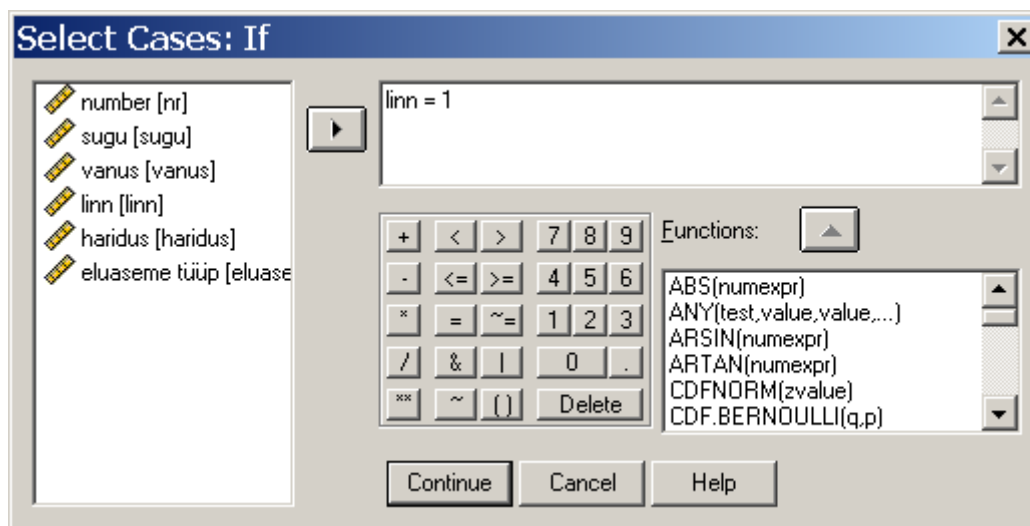
Tellimisakna alumises Output osas peab kindlasti olema valitud „filter out unselected cases“.

Sel juhul muudetakse mitteselekteeritud tunnused vastavate analüüside ajaks mitteaktiivseks. lejääänud variantide puhul salvestatakse selekteeritud objektid (inimesed) eraldi andmestikku või kustutatakse.



Pilt 4.

If nupu alt avanevas tellimisaknas (pilt 5) tuleb panna tingimuseks, et elukoha linnaks on Tallinn. Vastav kood andmestikus on „1“. Vajutada Continue ja OK.



Pilt 5.

8) Tehke Tallinna elanike kohta haridustaseme tunnuse sagedustabel.

Analyze – Descriptive Statistics – Frequencies

☐ Display filter.

Data – Select cases

Valida nupuasenditest „All cases“

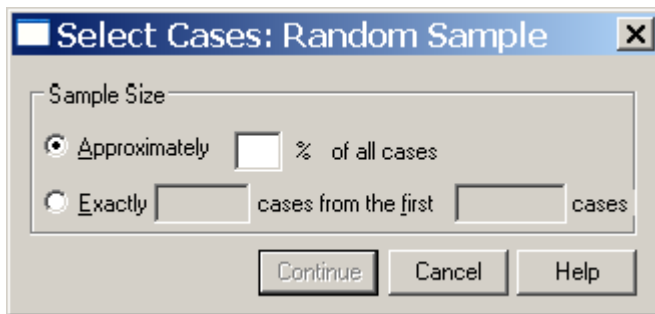
10) Selekteege filtri abiga andmestikust 5 juhuslikku inimest.

Data – Select cases

Valida nupuasenditest „Random sample of cases“

Nupu alt „Sample“ avanev tellimisaken on esitatud pildil 6. Selles tellimisaknas on võimalik valida juhuvalim protsendina kõigist objektidest või siis anda ette täpne indiviidide arv.

Praegu sobib teine variant „ exactly “, simesse lahtrisse sisestada soovitud inimeste arv (5) ja teise lahtrisse vastajate koguarv (20). Continue ja OK.



Pilt 6.

Seni kasutamata filtri võimalused on veel:

„ based on time or case range“, mis võimaldab valida objekte nende järjestuse alusel andmestikus.

„Use filter variable“, mille puhul saab kasutada mõnda olemasolevat filtertunnust.

Statistikapaketi SPSS kasutamine

Praktikum 3

Avage andmestik eestisooime.sav

Tutvuge tunnuste kirjeldustega
Utilities - Variables

oon t koo ta n and d

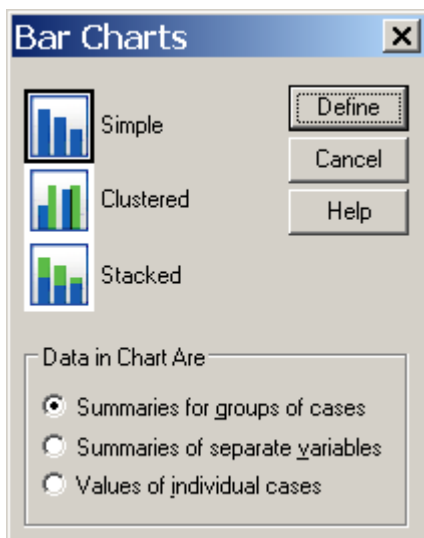
- Joonisel peab olema pealkiri ja allkiri
- Joonisel olevatele telgedele tuleb anda kirjeldused (nimi, ühikud)
- Vajadusel lisada joonisele legend
- Võrdluse aluseks olevad joonised võiksid olla sama tüüpi ja sama mastaabiga
- Jälgige, et joonisel oleks mõistlik skaala: mitte liiga lai ja mitte liiga kitsas, et joonise vaatajale ei jääks esitatud informatsioonist vale muljet.
- Vältige jooniseid, kus on esitatud liiga palju informatsiooni.

Tulpdiagrammid

a) tulpdiagramm t nn k

Graphs – Legacy Dialogs - Bar ... –

tage nupule Define.

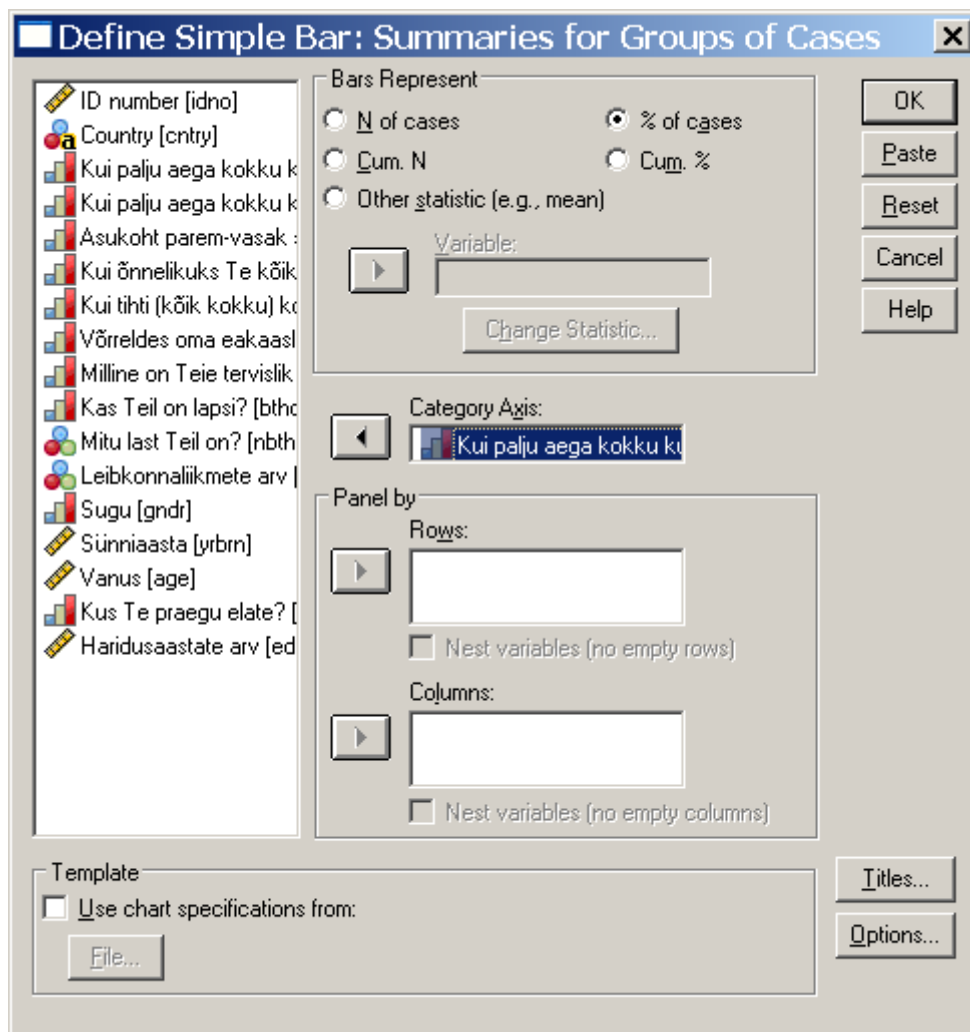


Pilt 1.

(vt pilt 2):

indiviidide arvu vastavas kategoorias/tulbas (N of cases); indiviidide protsendilist jaotust (% of cases); kumulatiivset indiviidide arvu (Cum. n of cases); kumulatiivset protsentjaotust

e.



Pilt 2.

1) Tehke tulpdiagramm tunnuse „Kui palju aega kokku kulutate televiisori vaatamisele tavalisel nädalapäeval “ kohta. Joonisel näidake protsente.

Graphs – Legacy Dialogs - Bar ... –

Selle joonise tellimisaken on näidatud ka joonisel 2.

Jooniste muutmine (redigeerimine): selleks teha valmis joonise kohal (väljundiaknas) vasaku hiireklahviga topeltklõps. Avaneb joonise redigeerimise aken, milles erinevatel joonise elementidel klikkides on võimalik nende elementide seadeid muuta.

Redigeerige joonis televiisorivaatamise aja kohta trükikõlbulikuks. Vajadusel muutke tulpade värve, lisage tulpadele sildikesed protsentidega, lisage samakõrgusjooned (grid).

2) Proovige sama joonis teha ka kahe teise jooniste tegemise käsu abiga

Graphs – Chart Builder

ja

Graphs – Interactive – Bar

3) Tehke joonis, kus on esitatud telerivaatamise gruppides vastajate keskmine vanus
Graphs – Legacy Dialogs - Bar ... – Simple

4) Tehke eelmises ülesandes nimetatud joonis estile ja Soomele võrdlevalt rea paneelides
Graphs – Legacy Dialogs - Bar ... – Simple
Tellimisaknas saata riigi tunnus „Panel by“ ja „Rows“ alla.
Lisage joonisel tulpadele keskmist vanust näitavad sildid.

t d a a ka t nn k
Graphs – Legacy Dialogs - Bar ... –

Category Axis: kategooria telg (x-telg)
Define Clusters by: tunnus, mille alusel jagatakse vastajad eri värvi tulpadesse.

5) Tehke joonis, kus on võrdlevalt esitatud Soome ja esti vastajate telekavaatamise protsentjaotus (s.t tunnuse „Kui palju aega kokku kulutate televiisori vaatamisele tavalisel nädalapäeval“ protsendid estis ja Soomes)
Graphs – Legacy Dialogs - Bar ... – Clustered (Summaries for groups of cases)
Joonisel esitage kindlasti protsendid

6) Võrrelge ka esti ja Soome vastajate raadiokuulamise ja ajalehtede lugemise harjumusi (tunnused „Kui palju aega kokku kulutate raadio kuulamisele tavalisel nädalapäeval“ ning „Kui palju aega kokku kulutate ajalehtede lugemisele tavalisel nädalapäeval“)

c) tulpdiagramm mitme tunnuse parameetrite d ko ta
Graphs – Legacy Dialogs - Bar ... –

7) Tehke joonis, kus oleks näidatud nende vastajate osakaal, kes vaatab televiisorit, kuulab raadiot või loeb ajalehte vastavalt vähemalt 0,5 tundi päevas. (S.o tunnuste kodeeritud väärtus on vähemalt 2, nupust „Change Statistic“ panna tingimuseks „Percentage above 1“).

8) Tehke sama joonis estile ja Soomele võrdlevalt
Graphs – Legacy Dialogs - Bar ... –

Joondiagrammid

Graphs– Legacy Dialogs – Line ...
Tellimine on sarnane tulpdiagrammide juures näidatuga.

□ Tehke joondiagramm vastajate haridusaastate arvu kohta.
Graphs– Legacy Dialogs – Line ... – Simple (Summaries for groups of cases)

10) Tehke joondiagramm vastajate haridusaastate arvu jaotuse kohta erinevates riikides.
Graphs– Legacy Dialogs – Line ... – Multiple (Summaries for groups of cases)

11) Tehke joondiaagramm vastajate haridusaastate arvu jaotuse kohta soo lõikes, joonised teha võrdlevalt (rea paneelis) riikide jaoks.

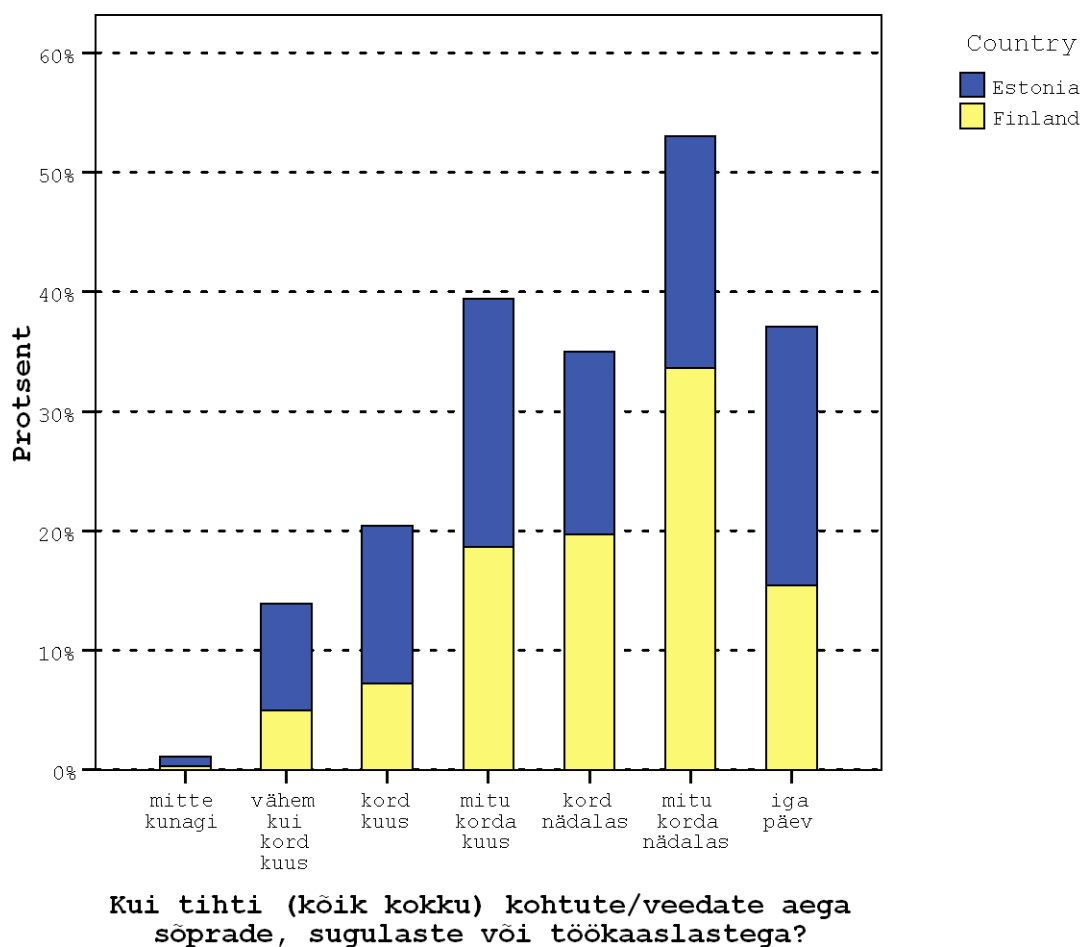
Graphs– Legacy Dialogs – Line ... – Multiple (Summaries for groups of cases)

Ringdiagrammid

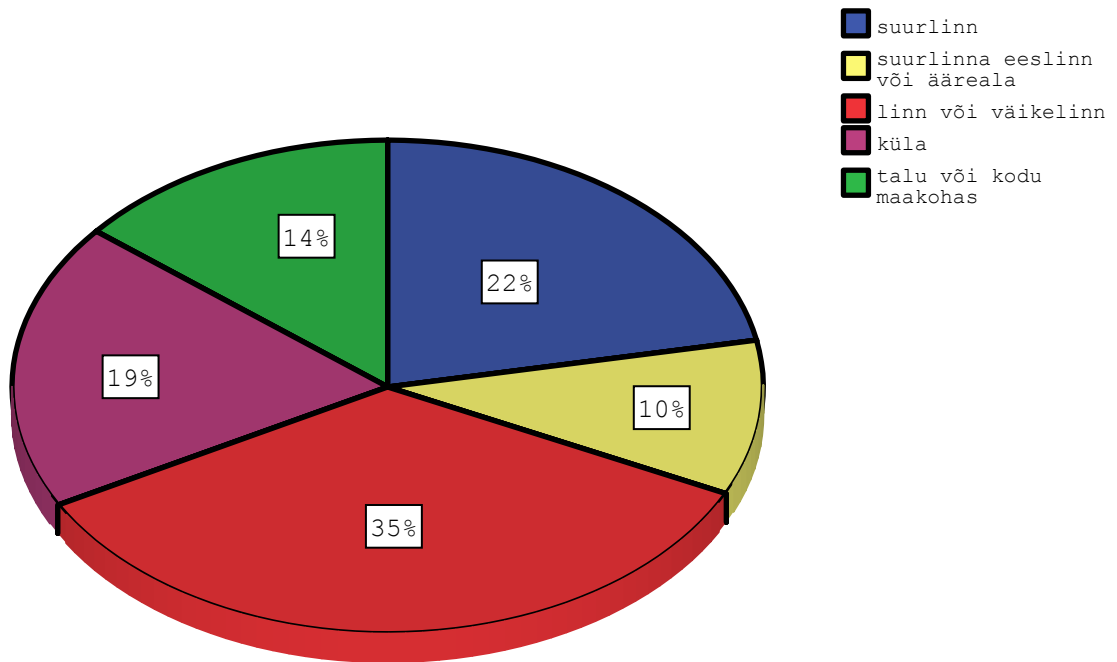
Graphs – Legacy Dialogs – Pie ... – Summaries for groups of cases

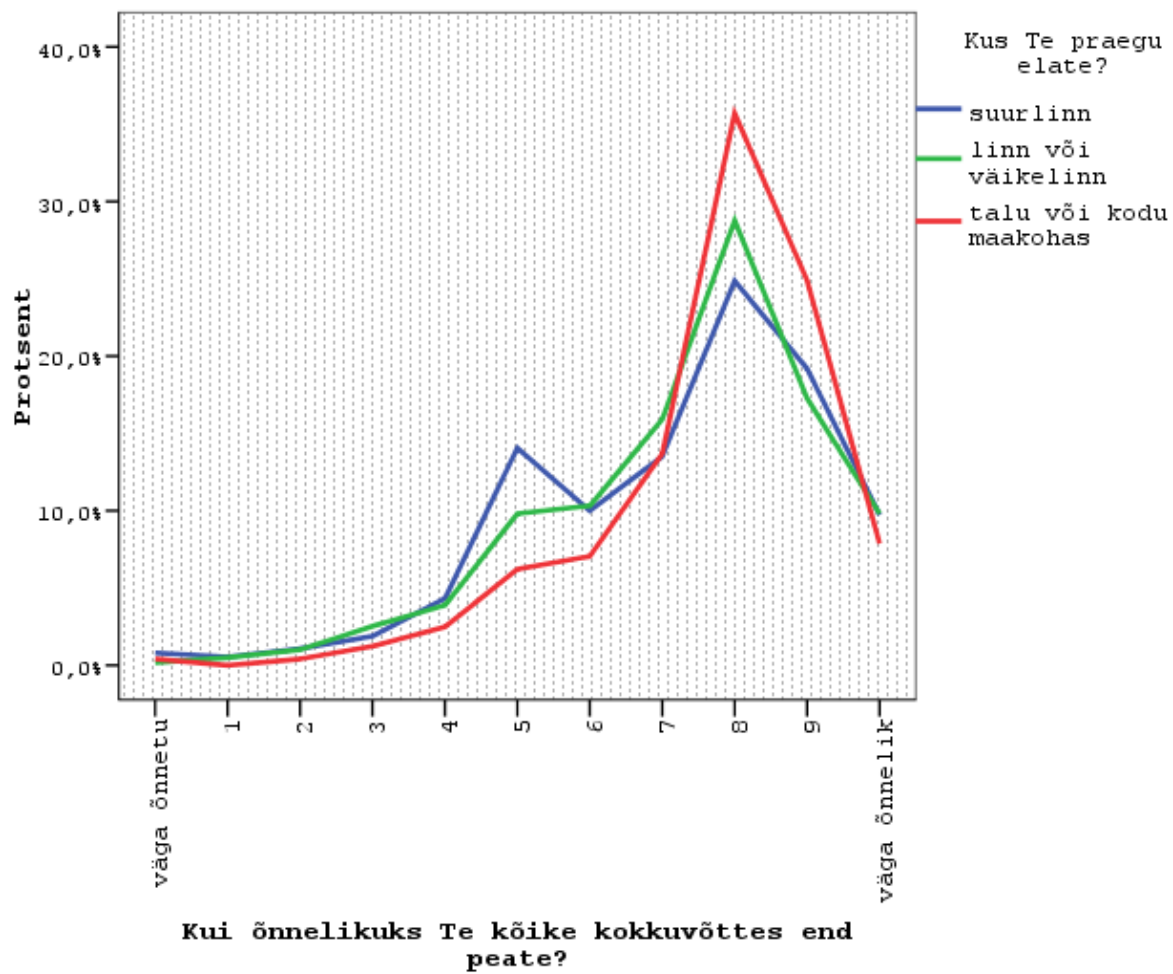
12) Tehke ringdiagramm vastajate hinnangute kohta oma tervislikule seisundile („ illine on Teie tervislik seisund “

13) Tehke järgi järgmised joonised



Kus Te praegu
elate?





Statistikapaketi SPSS kasutamine

Praktikum 4

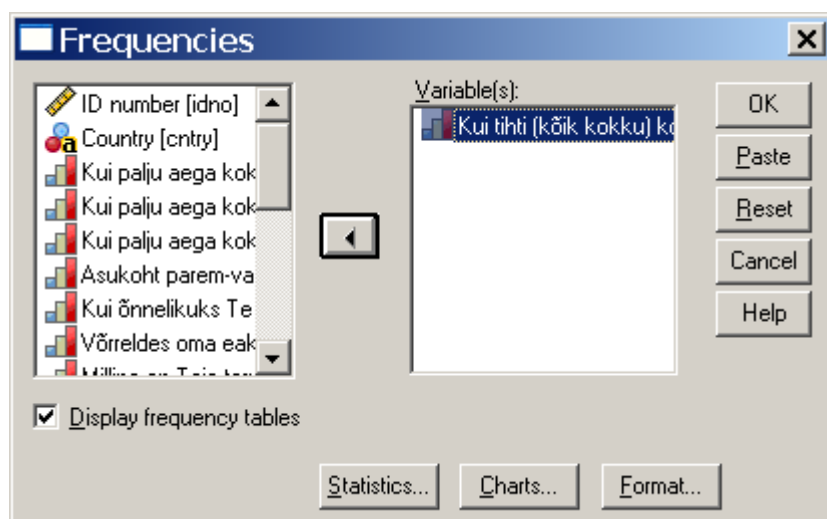
Avage andmestik eestisooime.sav
Tutvuge andmestikus olevate tunnuste kirjeldustega.

Sagedustabelid

Näide: sagedustabel tunnuse „Kui tihti (kõik kokku) kohtute veedate aega sõprade, sugulaste või töökaaslastega“ jaotuse kohta (vt ka pilt 1).

Selleks valige menüüdest:

Analyze – Descriptive Statistics – Frequencies



Valige tunnuste hulgast vajalik (Kui tihti) ning saatke see aknasse „Variable(s):“. Tunnuste nimekirja all peab olema sisse lülitatud „Display frequency tables“

Pilt 1.

Nupu alt „Statistics“ saab valida erinevaid statistilisi näitajaid (sellest pikemalt järgmises praktikumis).

Nupu alt „Charts“ on võimalik tellida lisaks sagedustabelile ka joonise vastava tunnuse jaotuse kohta.

Nupu alt „Format“ saab muuta sagedustabeli vormistust: tunnuse kategooriate järjestust tabelis, tabelite ja jooniste paigutust väljundiaknas jms.

Saadud tulemused koos tõlgendustega on esitatud tabelites 1 ja 2. esimene neist annab tunnuse üldinfo, tabel 2 on sagedustabel.

Statistics		
Kui tihti (kõik kokku) kohtute/veedate aega sõprade, sugulaste või töökaaslastega?		
N	Valid	1707
	Missing	6

Tabel 1.

Kui tellimisaknas pole „Statistics“ nupu alt tellitud ühtegi statistilist näitajat (nagu praeguses näites), siis esitatakse esimeses tabelis vastanud indiviidide arv (N Valid, antud juhul 1707) ning nende indiviidide arv, kellel antud küsimus oli vastamata ehk kellel esines puuduvaid väärtusi (N Missing, praegu 6).

Tabelis 2 antakse esimeses veerus esinenud tunnuse väärtuste või vastusevariantide loetelu. Teises veerus (Frequency) esitatakse vastava vastusevariandi andnud inimeste arv ning vastajate üldarvud.

Kolmandas veerus (Percent) on näidatud vastuste suhtelised sagedused, kui võtta arvesse kõik andmestikus leidunud indiviidid (ka need, kellel on antud tunnuse suhtes puuduvad väärtused).

Neljandas veerus (Valid Percent) esitatakse vastusevariantide osakaalud protsentides nende vastajate lõikes, kes on antud küsimusele vastanud. Ittevastajad (antud näites ka need, kes ütlesid „ei oska öelda“) on kõrvale jäetud. Samasti huvitab meid protsent küsimusele vastajatest, seega kasutatakse uurimistöös sagedamini „Valid Percent“ veerus näidatud protsente.

Viiendas veerus (Cumulative Percent) antakse kumulatiivsed suhtelised sagedused. Selleks liidetakse vastava rea ja kõigi varasemate ridade protsendid. Näiteks tabel 2 põhjal saab öelda, et 81,7% vastajatest kohtub oma sõprade, sugulaste või töökaaslastega harvemini kui iga päev ning 7,3% harvemini kui kord kuus.

Kui tihti (kõik kokku) kohtute/veedate aega sõprade, sugulaste või töökaaslastega

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	mitte kunagi	9	,5	,5	,5
	vähem kui kord kuus	116	6,8	6,8	7,3
	kord kuus	170	9,9	10,0	17,3
	mitu korda kuus	335	19,6	19,6	36,9
	kord nädalas	302	17,6	17,7	54,6
	mitu korda nädalas	463	27,0	27,1	81,7
	iga päev	312	18,2	18,3	100,0
	Total	1707	99,6	100,0	
Missing	eoö	6	,4		
Total		1713	100,0		

Tabel 2.

Topeltklõps tabelil väljundiaknas võimaldab tabelit ka redigeerida. Selleks saab kasutada menüüsid „Insert“, „Pivot“ ja „Format“. Samuti avanevad mõningad tabeli redigeerimise võimalused, kui klõpsata tabelile või selle elementidele parema hiireklahviga.

1) Tehke sagedustabel tunnuse „Kus Te praegu elate“ kohta.

Analyze – Descriptive Statistics – Frequencies

☐ Iitu protsenti vastajatest elab suurlinnas ☐ Illises asulatüübis elab kõige suurem osa vastajatest ☐ Iitu protsenti vastajatest elab külas, talus või maakohas

iged vastused: 22,0 linn või väikelinn (34,) 100 67,1 32,0

2) Tehke sagedustabel tunnuse „Kui tihti on Teie tervislik seisund“ kohta.

Analyze – Descriptive Statistics – Frequencies

☐ Iitu protsenti vastajatest oli hea või väga hea tervisega Kui palju oli vastajaid, kelle tervis on „rahuldav“

iged vastused: 56,6 621.

nn t kod n

Transform - Recode into Same Variables

Tunnuse ümberkodeerimine nii, et olemasolev tunnus asendatakse uuega.

Transform - Recode into Different Variables

Tunnuse ümberkodeerimine nii, et tekkiks eraldi uus ümberkodeeritud tunnus.

Näide: kodeerime ümber tunnuse „□ illine on Teie tervislik seisund□“ kolmeväärtuseliseks uueks tunnuseks „tervis“ koodidega:

1 – hea või väga hea

2 – rahuldav

3 – halb või väga halb

Transform - Recode into Different Variables

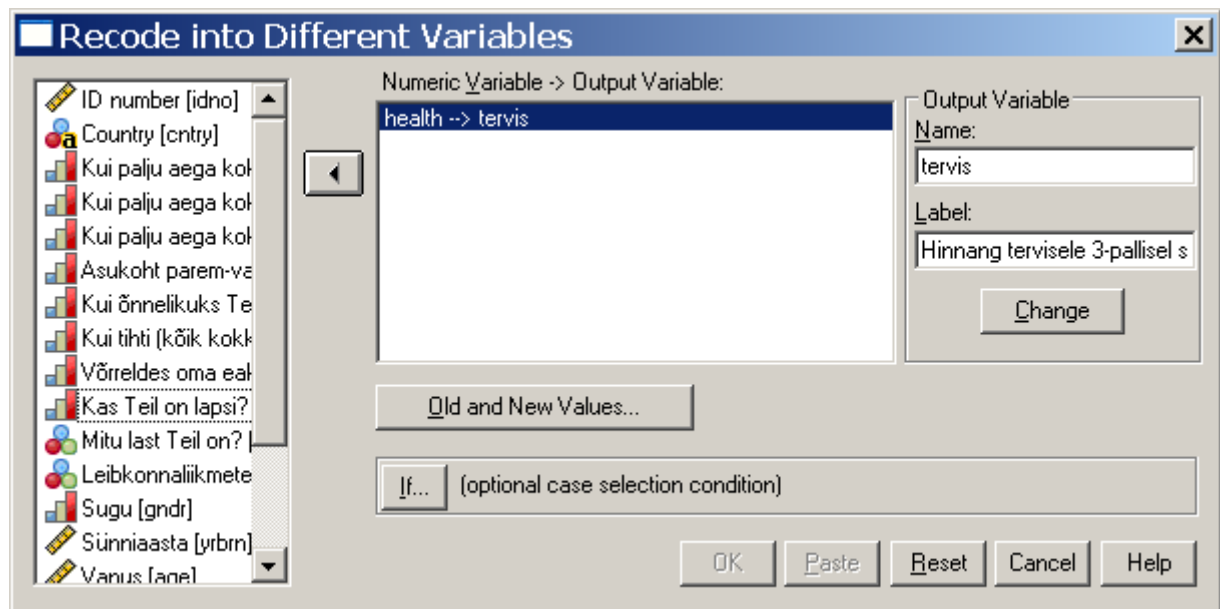
Avanevas tellimisaknas tuleks tunnuste nimekirjast valida ümberkodeerimisele minev tunnus (□ illine on Teie tervislik seisund) ja saata see parempoolsesse aknasse „Input Variable □□ Output Variable“. Oodustatavale uuele tunnusele tuleb anda ka nimi ja kirjeldus. Selleks on tellimisakna parempoolses ääres Output Variable all lahtrid:

Name: uue tunnuse lühinimi (kuni 8 tähte), antud näites saab selleks „tervis“

Label: uue tunnuse pikem kirjeldus. Antud näites on antud kirjelduseks „□ hinnang tervisele 3□ pallisel skaalal“

Tunnusele nime ja kirjelduse andmise järel vajutada nime kinnitamiseks nupule „Change“.

Seejärel peaks tellimisaken nägema välja selline nagu kujutatud pildil 2.



Pilt 2.

Järgnevalt tuleb määrata eeskiri tunnuse ümberkodeerimise jaoks. Selleks on vaja teada olemasoleva tunnuse (□ illine on Teie tervislik seisund□) kategooriate koodi. Antud juhul on tunnus kodeeritud selliselt: 1 – väga hea 2 – hea □ 3 – rahuldav □ 4 – halb □ 5 – väga halb. t saada soovitud 3□palline jaotus, on vaja omavahel liita kategooriad 1 ja 2 ning kategooriad 4 ja 5. Selleks tuleb valida nupp „Old and New Values“.

Nupu alt „Old and New Values“ avanev tellimisaken on jagatud kaheks osaks: vana ja uue tunnuse väärtusteks. Kõigile vana tunnuse (originaaltunnuse) väärtustele tuleb selles tellimisaknas omistada ka uue tunnuse väärtus.

Vana tunnuse väärtuste (Old Value) märkimiseks on järgmised võimalused (vt ka pilt 3):

Value: võimaldab märkide ühe üksikväärtuse (koodi).

System missing: võimaldab märkida kõik andmestikus esinevad puuduvad väärtused ehk tühjad kohad (missing values).

System or user missing: võimaldab märkida kõik andmestikus esinevad tühjad kohad ja ka kasutaja poolt defineeritud puuduvad väärtused.

Range: võimaldab märkida väärtuste (koodide) vahemiku. Sel juhul on vaja ette anda vahemiku kõige madalam ja kõige kõrgem väärtus (kood).

Range, LO ST through value: võimaldab märkida kõik väärtused alates kõige madalamast võimalikust kuni mingi etteantava piirini. Lahtrisse sisestatud väärtus haaratakse samuti kaasa.

Range, value through I ST: võimaldab märkida kõik mingist etteantud väärtusest kõrgemad koodid. Lahtrisse sisestatud väärtus haaratakse samuti kaasa.

All other values: kõik ülejäänud väärtused.

Uue tunnuse väärtuste (New Value) märkimiseks on kolm võimalust:

Value: üksikväärtus

System missing: puuduv väärtus (tühi koht andmestikus)

Copy old value(s): kopeeritakse vana tunnuse väärtus(ed)

Recode into Different Variables: Old and New Values

Old Value

☒ Value:

☐ System-missing

☐ System- or user-missing

☐ Range:

☐ Range, LOWEST through value:

☐ Range, value through HIGHEST:

☐ All other values

New Value

☒ Value:

☐ System-missing

☐ Copy old value(s)

Old --> New:

Add

Change

Remove

1 --> 1

2 --> 1

3 --> 2

4 --> 3

5 --> 3

☒ Output variables are strings Width: 8

☐ Convert numeric strings to numbers ('5'>5)

Continue Cancel Help

Pilt 3.

Kodeerides ümber tervisehinnangu tunnust, tuleb vanale väärtusele 1 omistada uus väärtus 1 vanale väärtusele 2 omistatakse uus väärtus 1 vanale väärtusele 3 uus väärtus 2 ning vanadele väärtustele 4 ja 5 omistatakse mõlemale uus väärtus 3. Iga sisestuse (muutuse) järel tuleb

vajutada nuppu „Add“. Juba lisatud ümberkodeerimise eeskirja saab muuta nupuga „Change“ ning kustutada nupuga „Remove“.

Kui kõigile vana tunnuse väärtustele on omistatud uued väärtused (nagu näidatud pildil 3), vajutada „Continue“.

Tunnuse ümberkodeerimisel on võimalik lisada ka tingimusi nupu alt „If“. Näiteks, et ümberkodeerimine teostataks vaid juhul, kui tegemist on Eesti elanikega, meestega vms. Seda võimalust vaatame lähemalt järgmistes praktikumides. Praegusel juhul vajutage lihtsalt tellimisaknas OK.

Uus ümberkodeeritud tunnus paigutatakse andmestikus kõige viimaseks.

Tekkinud tunnusele „tervis“ tuleb andmeakna „Variable View“ lehel panna ka kategooriate koodide kirjeldused (Values). Selle teostamiseks vaadake vajadusel 1. praktikumi juhendit.

4) Tehke sagedustabel Teie enda poolt ümberkodeeritud tervisehinnangu tunnuse kohta.
Analyze – Descriptive Statistics – Frequencies

Saadud tabel peaks välja nägema sarnane tabelile 3.

Hinnang tervisele 3-pallisel skaalal

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	hea	969	56,6	56,6	56,6
	rahuldav	621	36,3	36,3	92,9
	halb	122	7,1	7,1	100,0
	Total	1712	99,9	100,0	
Missing	System	1	,1		
Total		1713	100,0		

Tabel 3.

5) Kodeerige ümber tunnus „Kui palju aega kokku kulutate televiisori vaatamisele tavalisel nädalapäeval“ selliselt, et uuel tunnusel oleks järgmised grupid:

0 – üldse mitte

1 – kuni 1 tund (ärge arvestage siia neid inimesi, kes üldse televiisorit ei vaata)

2 – üle 1 tunni, kuni 2 tundi

3 – üle 2 tunni, kuni 3 tundi

4 – üle 3 tunni

6) Koostage sagedustabelid originaaltunnuse (Kui palju aega kokku kulutate televiisori vaatamisele tavalisel nädalapäeval) ja Teie poolt ümberkodeeritud tunnuse jaoks.

Statistikapaketi SPSS kasutamine

Praktikum 5

Erinevaid statistilisi näitajaid kirjeldava statistika puhul tuleks kasutada pidades silmas analüüsitavate tunnuste tüüpe: kas tegemist on nominaal-, järjestus- või arvtunnusega.

Statistilise näitaja kirjeldava statistika puhul (sulgudes ingliskeelne nimetus)

Nominaaltunnused

mood (mode) – kõige sagedasem väärtus või väärtusklass

mediaan (median)

mediaan (median) – punkt tunnuse skaalal, millest väiksemaid ja suuremaid väärtusi on variatsioonreas ühepalju. Mediaan jaotab skaala vaadeldava tunnuse seisukohalt kaheks võrdsagedaseks osaks.

kvantiilid (quantiles) – jagavad skaala mitmeks võrdsagedaseks osaks. Sagedamini kasutatavad kvantiilid on:

kvartiilid (quartiles) – jagavad tunnuse skaala neljaks võrdsagedaseks osaks. Näiteks alumine kvartiil on punkt, millest väiksemaid väärtusi on kogumis 25%. Ülemine kvartiil on punkt, millest suuremaid väärtusi on kogumis 25%.

detsiilid (deciles) – jagavad tunnuse skaala kümneks võrdsagedaseks osaks.

Arvtunnused

mood

mediaan

kvantiilid

aritmeetiline keskmine ehk keskvärtus (mean)

standardhälve (standard deviation) – tunnuse hajuvuse näitaja, mis mõõdab, kui kauged on keskmise vastaja keskmisest

dispersioon (variance) – standardhälbe ruut

Avage andmestik eestisooime.sav

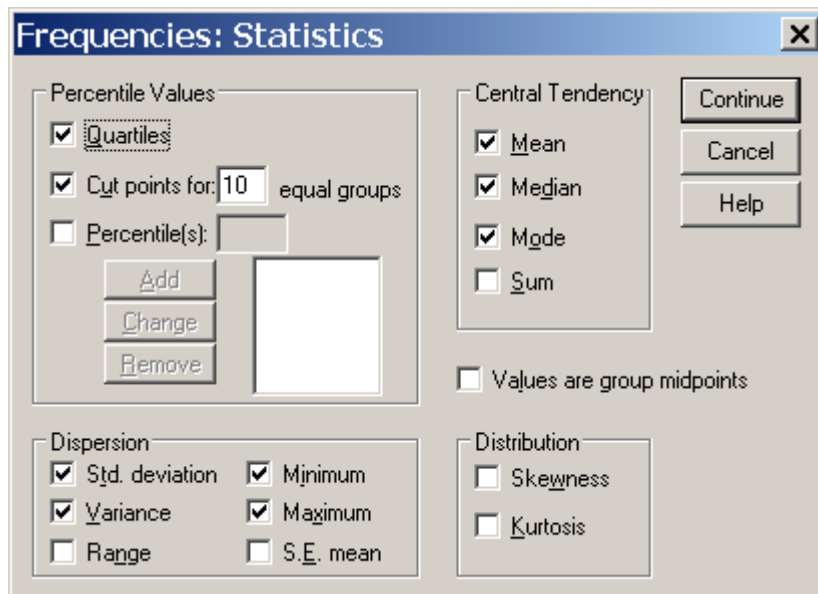
Kirjeldava statistika leidmiseks on SPSSis mitmeid erinevaid võimalusi. Esimesel korral vaadatuga on sarnane:

Analyse – Descriptive Statistics – Frequencies

Valida tunnused ning nupu alt „Statistics“ soovivad kirjeldava statistika näitajad.

1) Leidke vastajate keskmine vanus, selle kvartiilid ja detšiilid, mediaan, mood, standardhälve, dispersioon ning minimaalne ja maksimaalne väärtus.

Soovitus: kui olete statistilised näitajad valinud (nagu pildil 1), siis eemaldage tellimisaknas märged „Display frequency tables“



Pilt 1.

2) Tehke sagedustabel ning leidke keskmine hinnang küsimusele „Kui õnnelikuks Te kõike kokku võttes end peate “ („Display frequency tables“ alla jälle märke tagasi panna)
Tehke sagedustabelist ka joonis.

itu protsenti vastajatest andis enda õnnelikkusele hinnangu, mis oli üle üldise keskmise

Kirjeldava statistika tabel paigutatakse tellimisel alati sagedustabeli ette.

Analyze Descriptive Statistics Descriptives

Võimaldab tellida samuti kirjeldavat statistikat tunnuste kohta.

3) Leidke vastajate keskmine vanus, selle standardhälve ning minimaalne ja maksimaalne vastajate vanus käsuga

Analyze – Descriptive Statistics – Descriptives

Nupu alt „Options“ valida soovitud statistilised näitajad (pilt 2)



Pilt 2.

4) Leidke küsitlute seas lastega meeste keskmine vanus ja keskmine haridusaastate arv.

☐ltreerimiseks *Data – Select cases*

Keskmise leidmiseks

Analyze – Descriptive Statistics – Descriptives

5) Leidke alla 30-aastaste lastega naiste keskmine laste arv.

Kui suur on maksimaalne laste arv alla 30-aastaste naiste seas (s.t kui palju lapsi on naisel, kellel antud kogumist on kõige rohkem lapsi)

Analyze – Descriptive Statistics – Frequencies

k t d

Analyze – Descriptive Statistics – Explore

Dependent list: sõltuv tunnus (see, mille keskmist on vaja)

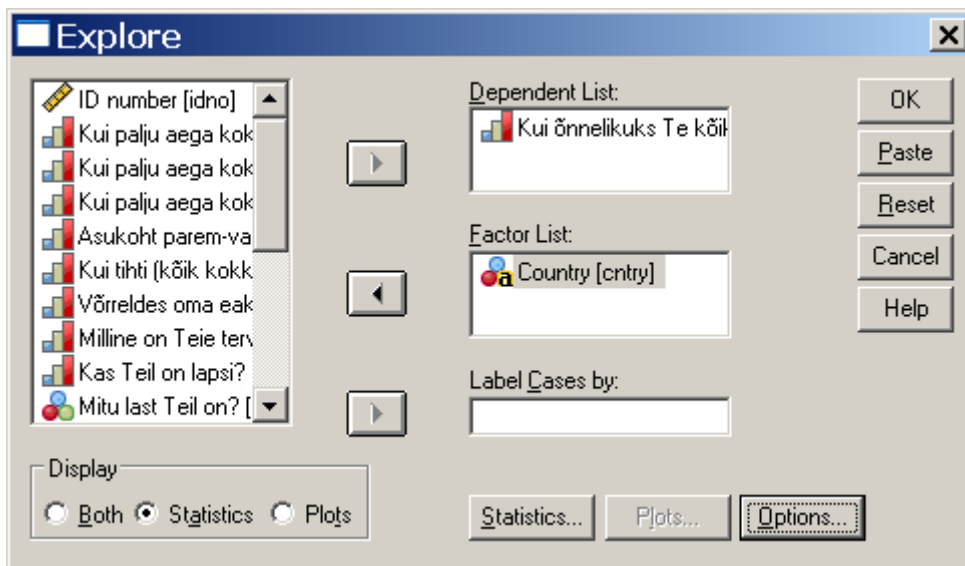
☐actor list: grupeeriv tunnus

Tunnuste nimekirja alt valida ainult „Statistics“ (ja mitte jooniseid)

„Options“ nupu alt võib soovi korral reguleerida seda, kuidas käsitletakse puuduvaid väärtusi

„Statistics“ nupu alt saab muuta usalduspiiride laiust.

Vt ka pilt 3



Pilt 3.

6) Leidke vastajate keskmised hinnangud enda õnnelikkusele riigi lõikes.

Analyze – Descriptive Statistics – Explore

Dependent list: Kui õnnelikuks

☐actor list: Country

7) Leidke vastajate keskmised hinnangud enda õnnelikkusele sõltuvalt nende tervislikust seisundist.

8) Kontrollige, kas vastavad tõele järgmised väited (kasutage vajadusel käsku *Data – Split file*), põhjendage oma vastuseid arvuliselt:

a) Soomes on keskmine laste arv suurem kui ☐estis (kui vaadata neid inimesi, kellel üldse on lapsi)

b) Soomes on 4 ja enama lapsega vastajaid rohkem kui ☐estis

- c) Nende inimeste hulgas, kelle tervislik seisund on hea või väga hea, on naised protsentuaalselt vähem kui mehed.
- d) laste seas, kelle tervislik seisund on hea või väga hea, on mehi protsentuaalselt rohkem kui naisi.
- e) lastis on nende inimeste keskmine vanus, kelle tervislik seisund on hea või väga hea, kõrgem kui Soomes
- f) lastis on õnnelikkuse tunnuse (Kui õnnelikuks Te kõike kokku võttes end peate) hajuvus suurem kui Soomes

Statistikapaketi SPSS kasutamine

Praktikum 6

Avage andmestik gss□1.sav

Tutvuge tunnuste kirjeldustega

Utilities - Variables

Või sellise nupu alt



äärake tunnuste tüüpe (nominaaltunnus, järjestustunnus, arvtunnus)

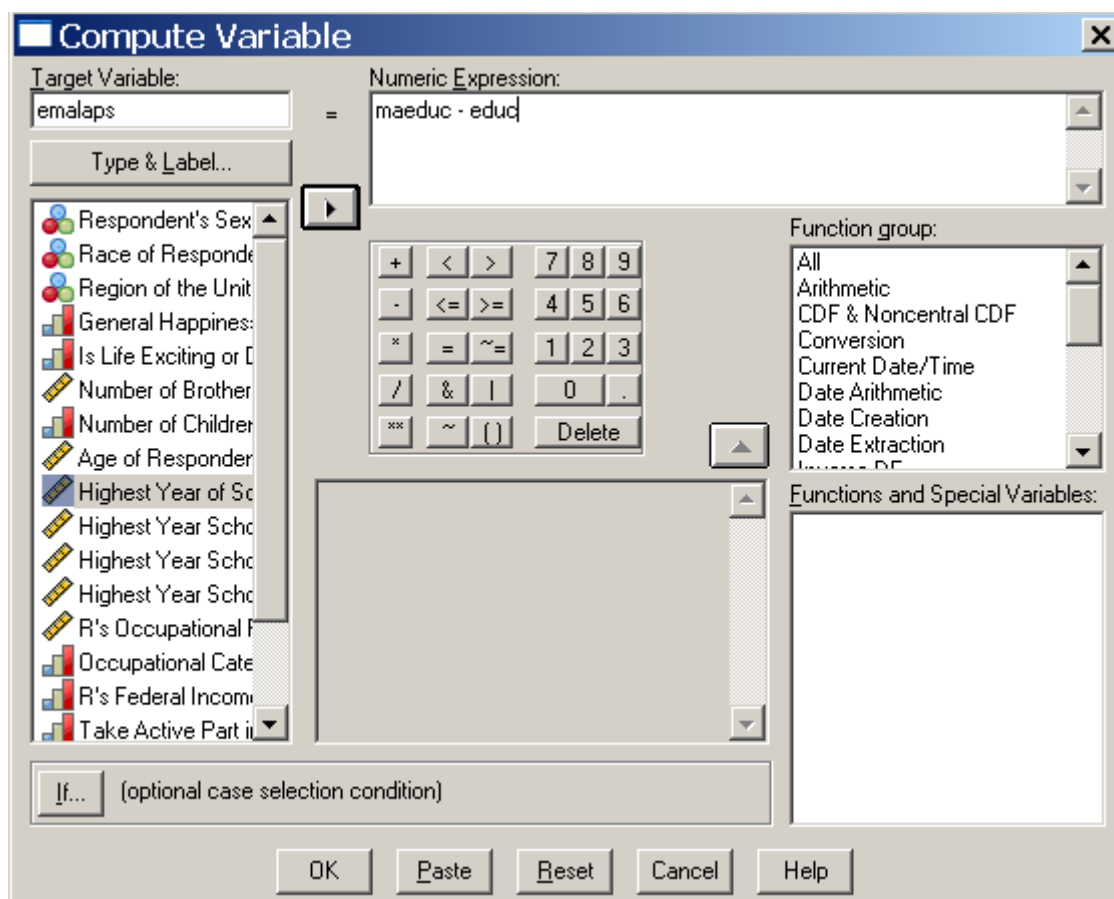
Uute tunnuste moodustamine

enüü *Transform*

Transform – Compute Variable

Uue tunnuse moodustamine

- 1) oodustage uus tunnus „ema ja lapse haridusaastate vahe“ (vt ka pilt 1)



Pilt 1.

Transform – Compute Variable

Target Variable: uue tunnuse nimi, näiteks „emalaps“

Type ☐ Label: uue tunnuse tüüp ja kirjeldus (ema ja vastaja haridusaastate vahe)

Numeric ☐ Expression: uue tunnuse arvutamise valem

maeduc \backslash educ

function group: sealt saaks valida mitmesuguseid funktsioone (näiteks summa, keskmine jne)

If: sealt saaks panna tingimuse, mille puhul arvutus teha (praegu pole seda vaja)

Vaadake uue tekkinud tunnuse jaotust (Analyse – Descriptive Statistics – Frequencies)

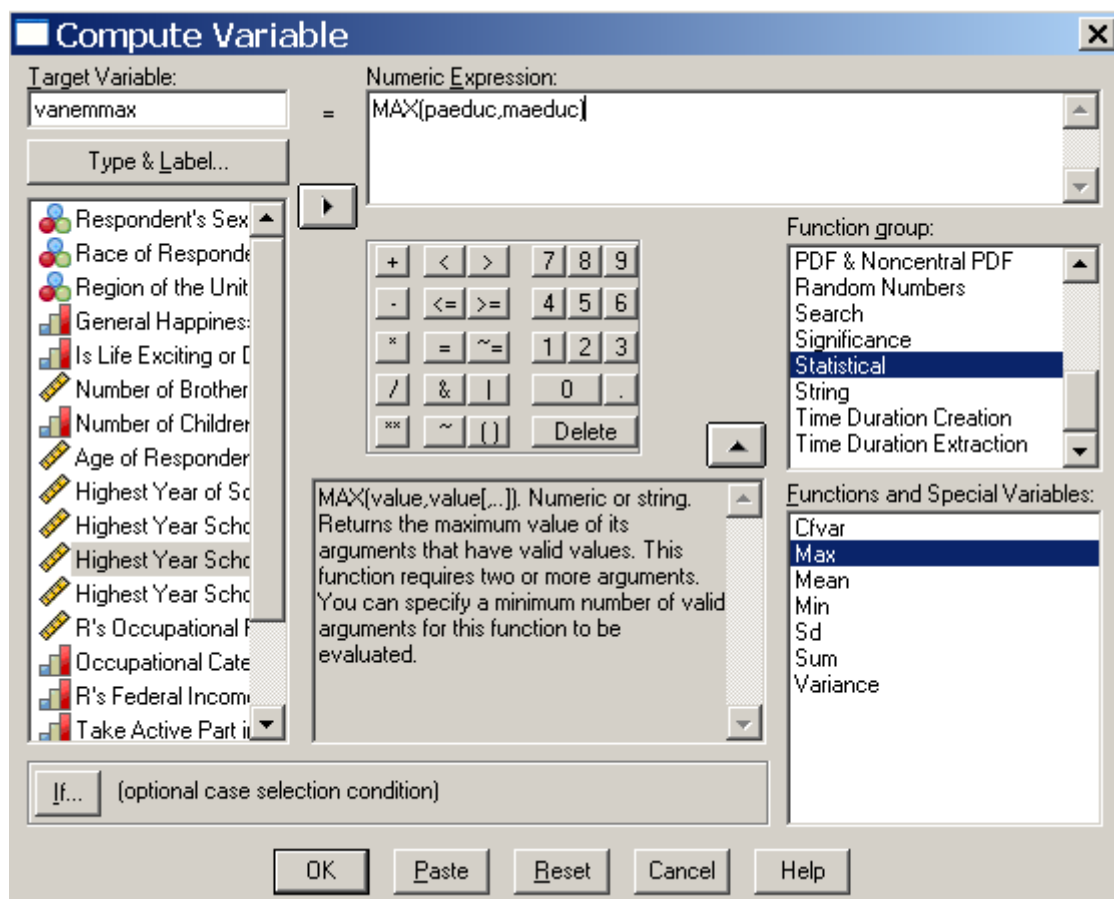
2) Oodustage uus tunnus, mis näitab vanemate maksimaalset haridusaastate arvu (s.t on haridusaastate arv sellel vanemal, kellel on kõige kõrgem haridustase)

Transform – Compute Variable

Target Variable: vanemmax

Numeric ☐ Expression

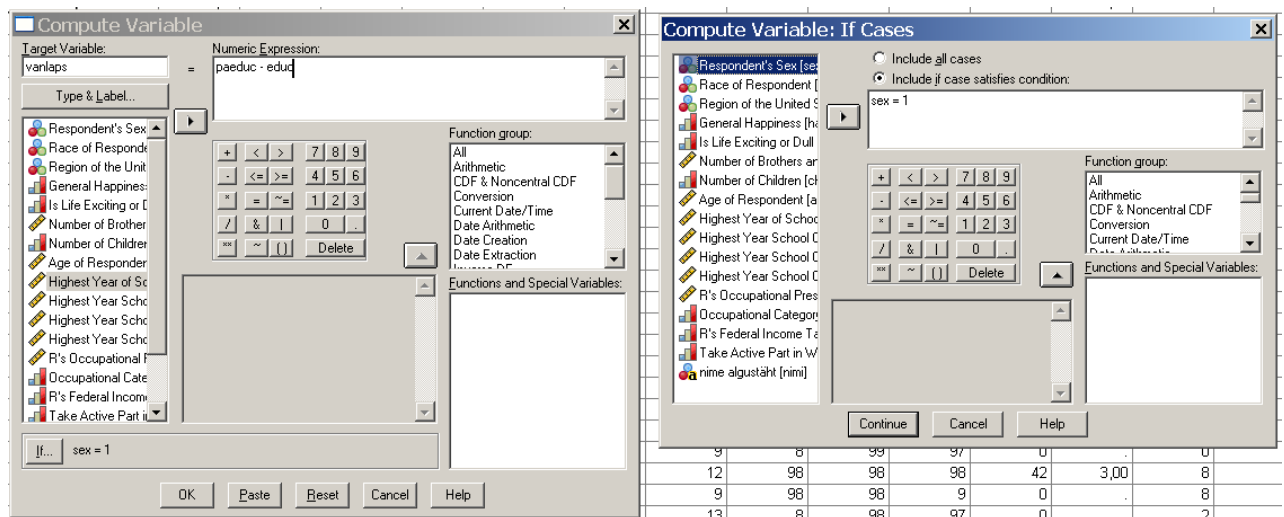
Valige funktsioonidest „Statistical“ ning statistiliste funktsioonide seast maksimum (Max) ning küsimärkide asemele saatke vanemate haridusaastate arvu tunnused (pilt 2).



Pilt 2.

3) Oodustage uus tunnus „vanema ja lapse haridusaastate vahe“, mis meeste puhul oleks „isa ja lapse haridusaastate vahe“ ning naiste puhul „ema ja lapse haridusaastate vahe“.

Selleks tuleb tunnus defineerida 2 korda, esmalt meestele (nupu alt If panna tingimuseks, et „Sex = 1“ kui teete tunnust meeste kohta ning „Sex = 2“ kui teete tunnust naiste kohta). Kõigete kohta käiva osa lahendus on näidatud pildil 3.



Pilt 3.

4) oodustage andmestikus olemasolevate tunnuste põhjal veel üks uus tunnus, arvutuseeskiri mõelge ise.

t t n n t ood ta n t t o n d a t

Transform – Count Values within Cases

Loeb kokku ühe objekti puhul vastava väärtusega tunnuste arvu

Näiteks: andmestikus on 4 tunnust haridusaastate arvu kohta (vastaja, tema ema, tema isa ja tema abikaasa haridusaastate arv). Lugege kokku vähemalt 12 haridusaastaga tunnuste arv indiviidil (vt ka pilt 4).

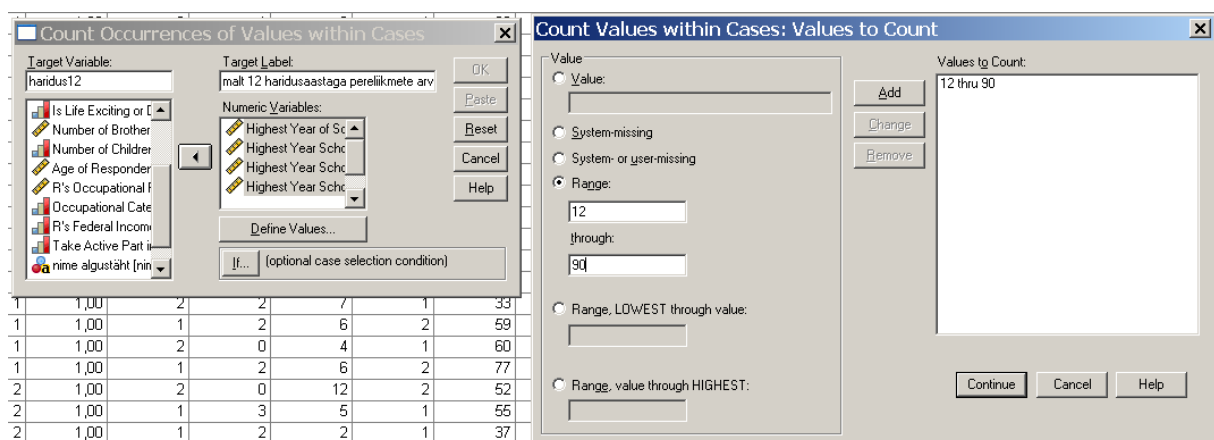
Target Variable: uue tunnuse nimi (näiteks haridus12)

Target Label: uue tunnuse kirjeldus (näiteks „Vähemalt 12 haridusaastaga pereliikmete arv“)

Numeric Variables: tunnused, mille väärtusi loetakse (antud juhul 4 haridusaastate arvu tunnust)

Define Values: määrata tingimus, millele vastavad väärtused kokku loetakse (antud juhul sobiks näiteks „12 kuni 0“, et mitte haarata kaasa puuduvaid väärtusi)

If: saab panna tingimuse, millisel juhul kokkulugemist teha



Pilt 4.

5) oodustage tunnus selle kohta, kas vastaja on 30 aastat vana

- 6) □ oodustage tunnus selle kohta, kas vastaja on vanuses 30–40 aastat.
7) □ oodustage tunnus selle kohta, kas vastaja on vähemalt 12 aastat koolis käinud sellistele vastajatele, kes on vähemalt 60 aastat vanad.

nn t kod n

Transform - Recode into Same Variables

Tunnuse ümberkodeerimine nii, et olemasolev tunnus asendatakse uuega.

Transform - Recode into Different Variables

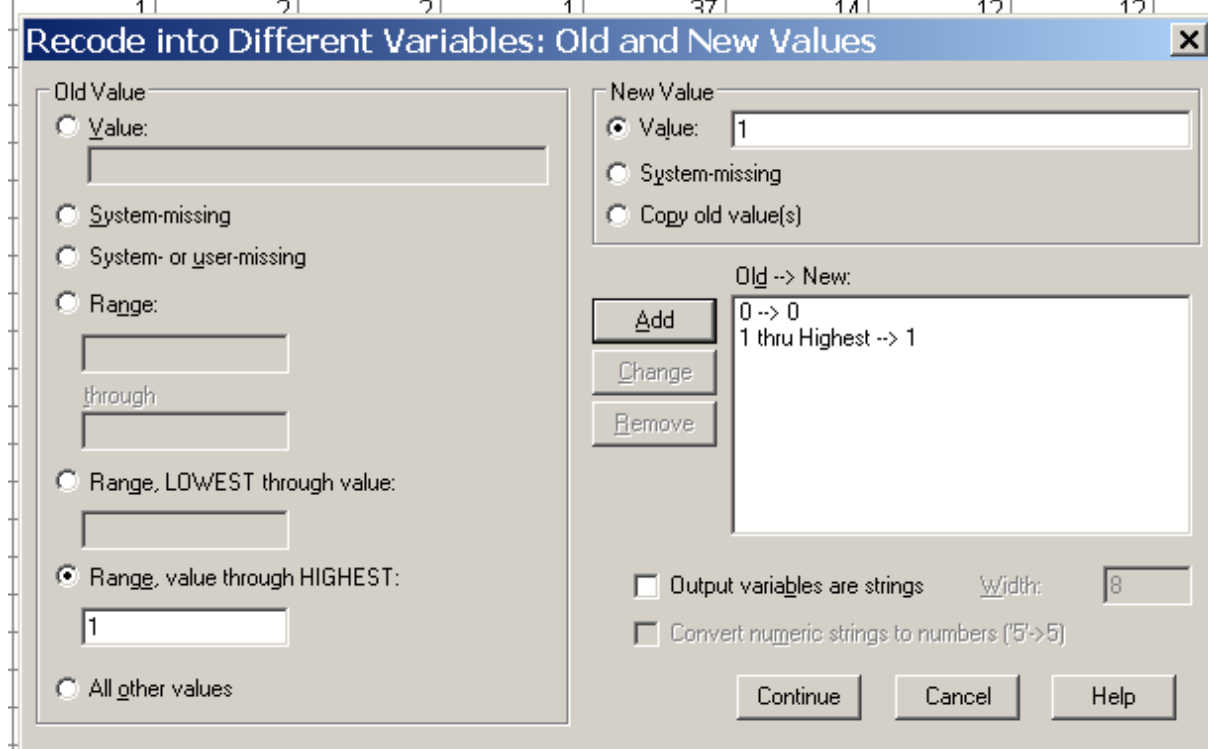
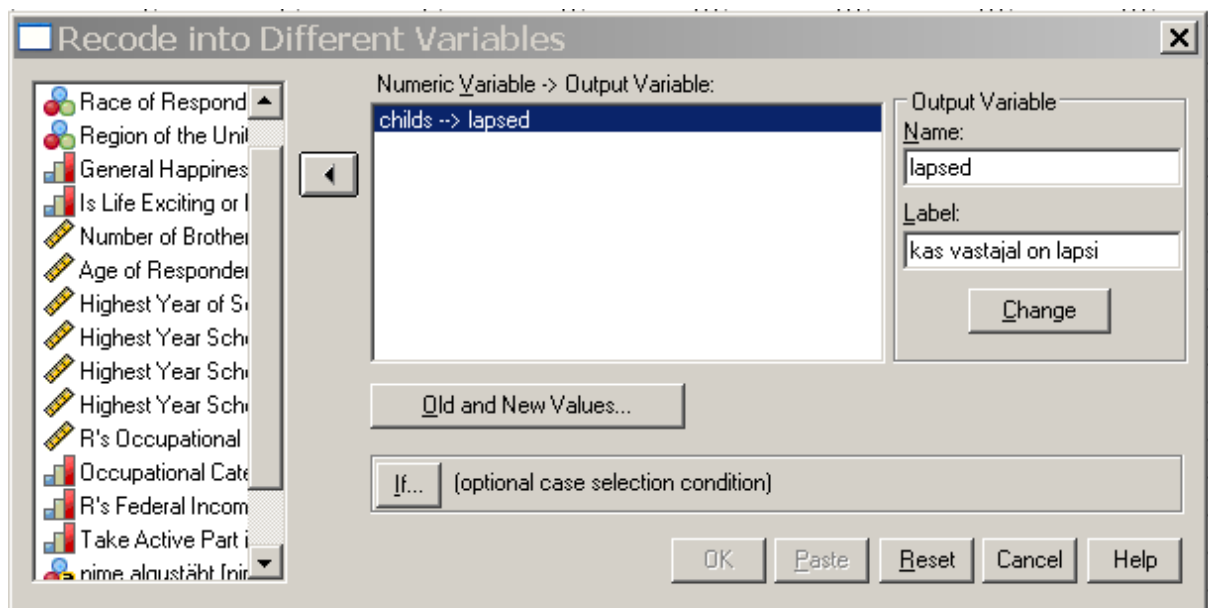
Tunnuse ümberkodeerimine nii, et tekkiks eraldi uus ümberkodeeritud tunnus.

- 8) Kodeerige vastaja laste arvu tunnus „Number of Children“ ümber kaheväärtuseliseks selliselt, et see näitaks, kas vastajal üldse on lapsi (väärtused 1 või rohkem) või ei ole (väärtus 0).

Transform - Recode into Different Variables

Vajadusel vaadake eelmiste praktikumide juhendeid ning pilti 5.

Kontrollige andmestikust järgi, kas ümberkodeerimine läks õigesti.



Pilt 5.

Transform - Automatic Recode

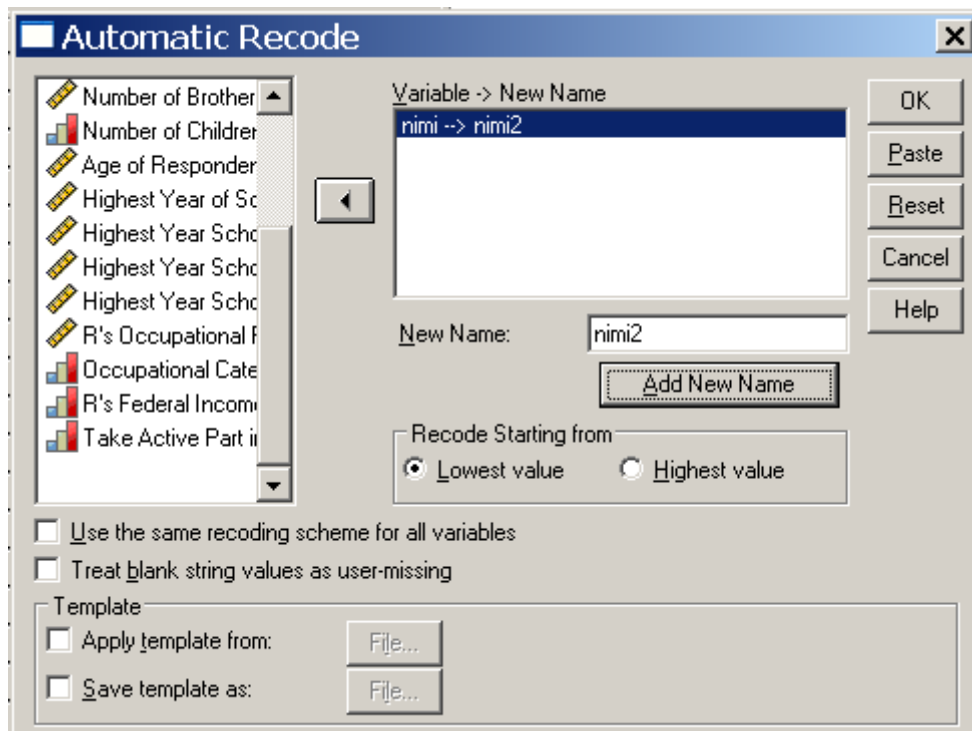
Kodeerib sõnalised tunnused (String) ümber arvulisteks

☐ Kodeerige tunnus „nimi“ ümber arvulisteks tunnuseks

Transform - Automatic Recode

Valige vana tunnus, pange uuele tunnusele uus nimi ning vajutage kindlasti ka „Add New Name“

Vt ka pilt 6.



Pilt 6.

Transform - Visual Binning

Arvuliste tunnuste ümberkodeerimine gruppideks.

10) Kodeerige andmestikus olev vanuse tunnus ümber vanusegruppideks järgnevalt:

1 – kuni 30a

2 – 31–40a

3 – 40–60a

4 – 61a ja vanem

Transform - Visual Binning

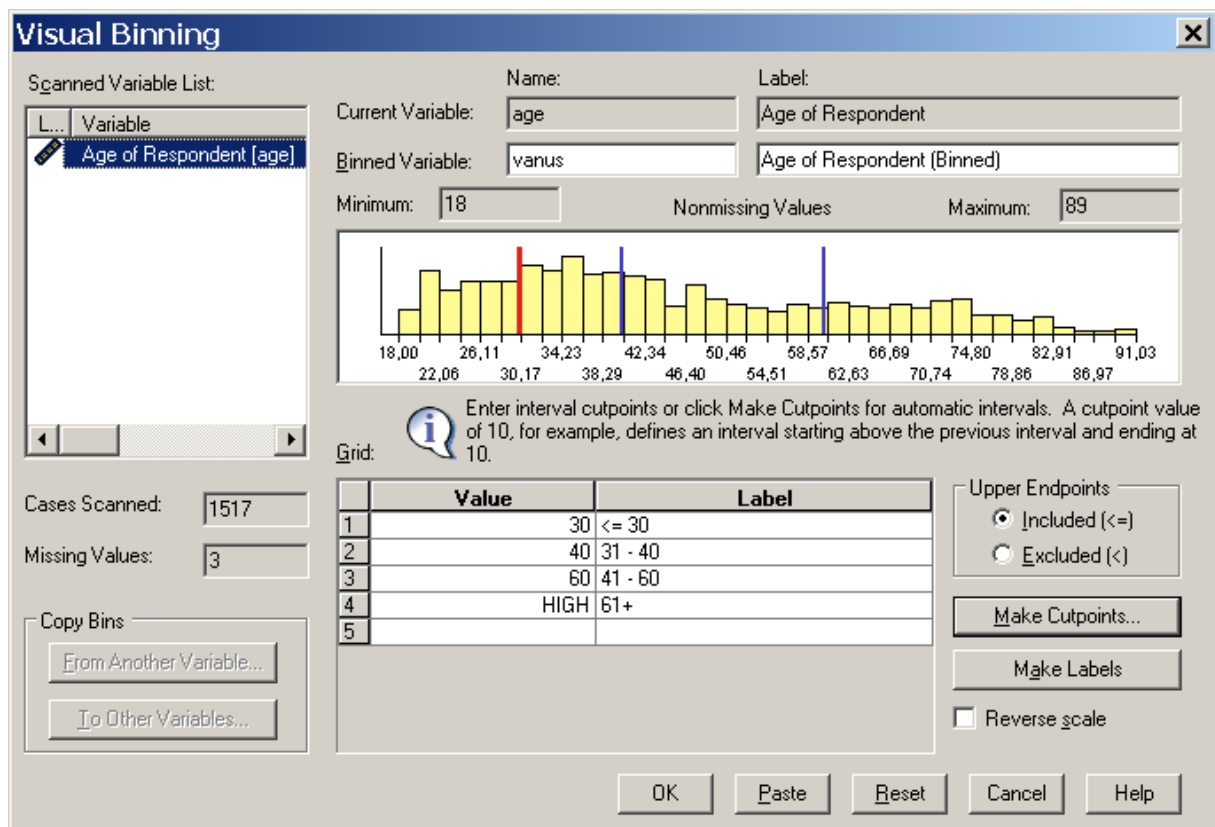
Valida tunnus (Age of Respondent)

Uues aknas muuta tunnus aktiivseks ja panna sellele uus nimi („inned Variable“ juures „Name“), näiteks „vanus“

äärata gruppide piirid.

„ake labels“ tekitab grupile kirjelduse.

igesti täidetud tellimisaken on näidatud pildil 7.



Pilt 7.

11) Kodeerige ümber vastaja haridusaastate arvu tunnus Teile mõistlikuna tunduval viisil.
Transform - Visual Binning

Transform - Rank Cases

oodustatakse uus tunnus, mille väärtuseks on indiviidi järjekorranumber mingi teise tunnuse väärtuste alusel.

Variable: tunnus, mille alusel indiviidide järjekord moodustatakse

☐ y: grupeeriv tunnus

Rank Types:

Rank ☐ järjekorranumber ehk astak

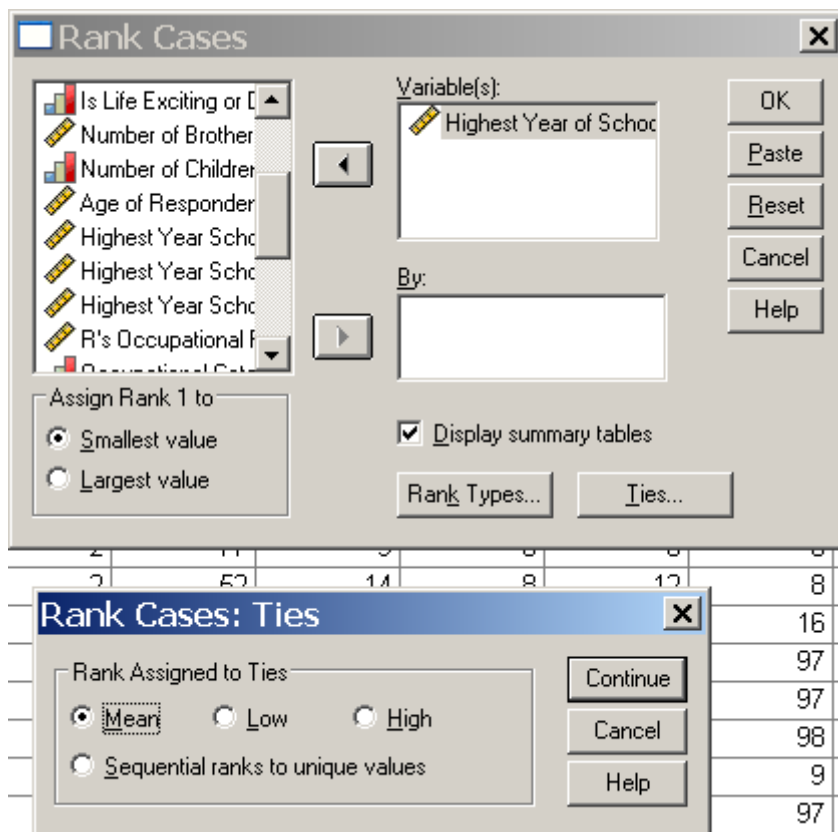
Ntiles ☐ kvantiilid

Ties: juhised selliseks olukorraks, kui mitmel indiviidil on tunnuse väärtus sama

13) oodustage tunnus, mille väärtuseks oleks vastaja järjekorranumber (astak) tema haridusaastate arvu alusel

Transform - Rank Cases

igesti täidetud tellimisaken selle ülesande puhul on näidatud pildil 8.



Pilt 8.

- 14) oodustage tunnus, mille väärtuseks oleks vastaja vanusekvartiil.
- 15) oodustage tunnus, mille väärtuseks oleks vastaja järjekorranumber haridusaastate arvu alusel temaga samast soost inimeste seas.

Statistikapaketi SPSS kasutamine

Praktikum 7. Keskmiste ja jaotusparameetrite d

Avage andmestik keskmised.sav
Tutvuge tunnuste kirjeldustega

k t a aot a a t t d n t t nn d

Käsuga

Analyze – Compare Means – Means

on võimalik võrrelda ühe tunnuse parameetreid teise tunnuse gruppides. Tulemuseks antakse tabel soovitud näitajatega gruppides ning ka tulemused kõigi gruppide kohta kokku.

Analyze – Compare Means – Means

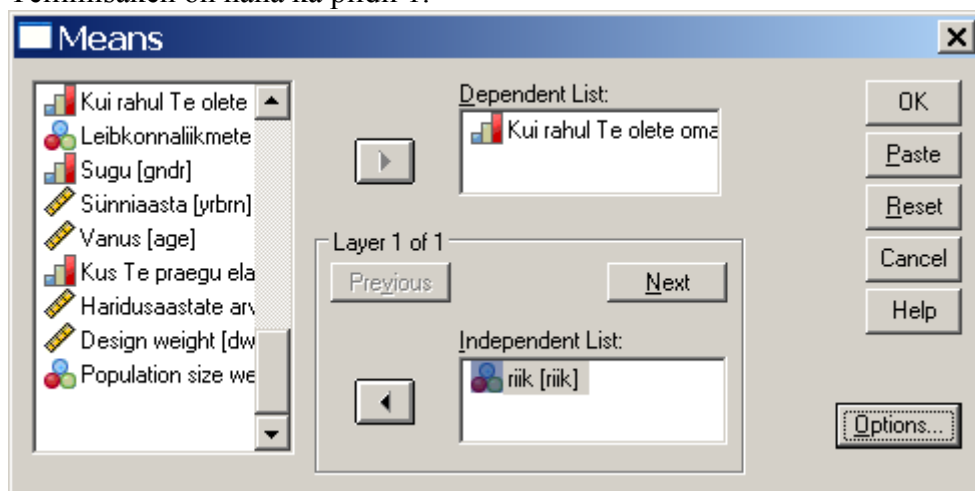
Dependent List: sõltuvad tunnused (need, mille keskmist tahame saada)

Independent List: grupeerivad tunnused

Layer: erineva taseme grupeerivad tunnused (mitu grupeerivat tunnust üksteise sees)

Options: tabeli lahtrites näidatava valik

Tellimisaken on näha ka pildil 1.



Pilt 1.

1) Leidke tunnuse „Kui rahul Te olete oma senise eluga“ keskmine väärtus ja standardhälve erinevates riikides.

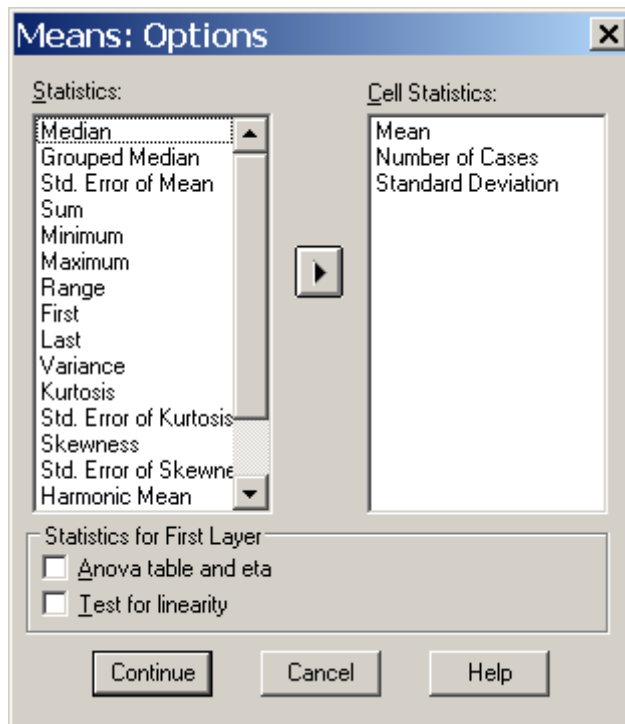
Analyze – Compare Means – Means

Dependent List: Kui rahul Te olete oma senise eluga

Independent List: riik

Options: automaatselt on sisse lülitatud keskmine, standardhälve ja vastajate arv

Options menüü alt automaatselt avanev vaade on esitatud pildil 2.



Pilt 2.

Tulemuste esitus n d

Keskmine hinnang rahulolule senise eluga oli kõrge (keskmine 7.1, standardhälve 1.8 punkti)

Teksti sees kasutamiseks sobiks ka

($m=7.1$, $sh=1.8$)

($\bar{x}=7.1$, $SD=1.8$)

Keskmine hinnang oli 7.1 punkti (standardhälve 1.8).

2) Leidke vastajate keskmised hinnangud küsimusele „Kui rahul Te olete oma praeguse materiaalse olukorraga“ riikide lõikes.

☐ illega on vastajad keskmiselt rohkem rahul, kas eluga tervikuna või oma materiaalse olukorraga ☐ Kummal tunnusel on suurem hajuvus ☐

Kas keskmiste erinevuse muster on sarnane kõigis riikides illises riigis on erinevused nende kahe tunnuse keskmistes kõige suuremad

3) Leidke vastajate keskmised hinnangud küsimustele „Kui rahul Te olete oma senise eluga“ ja „Kui rahul Te olete oma praeguse materiaalse olukorraga“ riigi ja vastaja soo lõikes. Selgitage välja, kas erinevates riikides on soolisi erinevusi rahulolus eluga ja materiaalse olukorraga. Sõnastage järeldused.

Analyze – Compare Means – Means

Layer: Next: lisada tunnus „Sugu“

Keskmine ja selle usalduspiirid

Analyze – Descriptive Statistics – Explore

Dependent list: sõltuv tunnus (see, mille keskmist on vaja)

☐ factor list: grupeeriv tunnus (pole tingimata vajalik)

Tunnuste nimekirja alt valida ainult „Statistics“ (ja mitte jooniseid)

„Options“ nupu alt võib soovi korral reguleerida seda, kuidas käsitletakse puuduvaid väärtusi

„Statistics“ nupu alt saab muuta usalduspiiride laiust.

4) Leidke tunnuse „Kui rahul Te olete oma senise eluga“ keskmine väärtus ja keskmise 5 usalduspiirid.

Analyze – Descriptive Statistics – Explore

Dependent list: Kui rahul Te olete oma senise eluga

Tulemuste tabelis usalduspiirid on näidatud

5 Confidence Interval for ean

Lower bound – alumine

Upper bound – ülemine

5) Leidke tunnuse „Kui rahul Te olete oma senise eluga“ keskmine väärtus ja keskmise usalduspiirid erinevates riikides

Dependent list: Kui rahul Te olete oma senise eluga

Factor list: riik

Nupu alt „Statistics“ märkida usaldusvahemiku laiuseks .

Joonised usalduspiiride kohta

Graphs – Legacy Dialogs – Error Bar

6) Tehke joonis tunnuse „Kui rahul Te olete oma senise eluga“ keskmiste ja nende 5 usalduspiiride kohta eri riikides.

Graphs – Legacy Dialogs – Error Bar

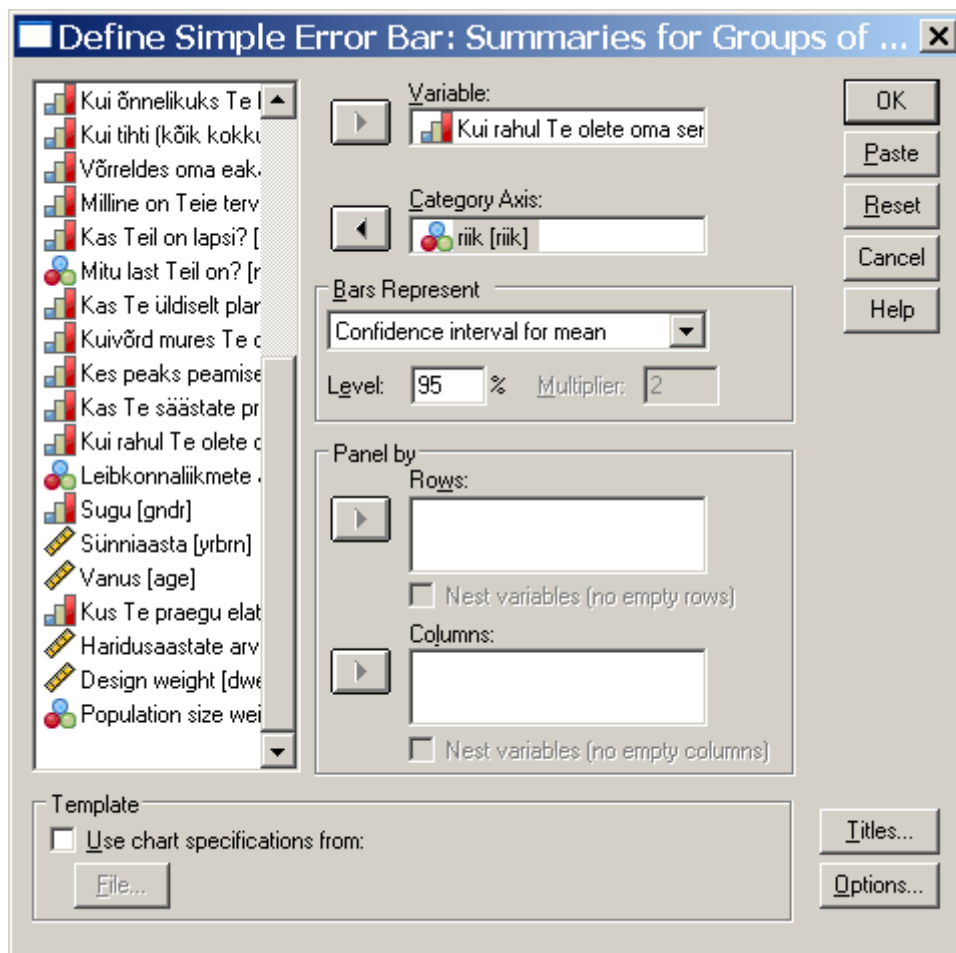
Simple (Summaries for groups of cases)

Variable: Kui rahul Te olete oma senise eluga

Category Axis: Riik

Tellimisaken on näidatud pildil 3.

Tulemuseks saadud joonisel märgib keskmist ring ning usaldusvahemiku laiust näitavad selle ümber olevad haarad.



Pilt 3.

7) Tehke joonis andmestikus olevate usalduse tunnuste (alates „Usaldus: parlament“ kuni „Usaldus: RO“) keskmiste väärtuste ja keskmiste 0 usalduspiiride illustreerimiseks

Graphs – Legacy Dialogs – Error Bar

Simple (Summaries of separate variables)

8) Tehke ülesandes 7 antud joonis selliselt, et joonisel oleks usalduse tunnuste keskmised ja standardhälbed. Selleks tellimisaknas „ars represent“ alt valida „Standard deviation“ (ning „multiplier“ võiks olla 1)

Statistikapaketi SPSS kasutamine

Praktikum 8 k t d : T-t t a d oonana

Avage andmestik keskmised.sav

Tutvuge tunnuste kirjeldustega

T-test

Analyze – Compare Means – Independent-Samples T Test

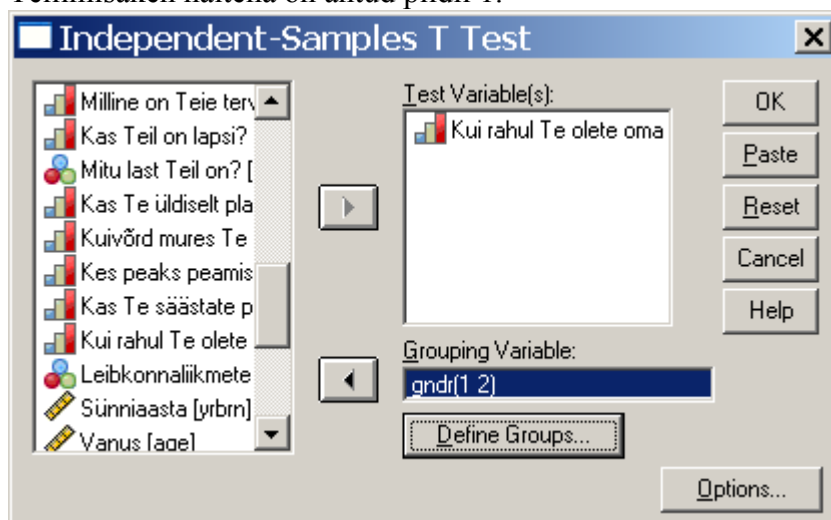
Test Variable(s): sõltuv tunnus (see, mille keskmist tahame saada), võib olla ka mitu

☐ Grouping Variable: grupeeriv tunnus

Define Groups: määrata võrreldava kahe grupi koodid andmestikus

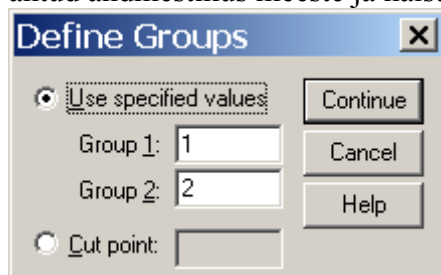
Options: saab määrata, kuidas käitutakse puuduvate väärtustega (ning keskmiste erinevusele antud usaldusvahemiku laiuse)

Tellimisaken näitena on antud pildil 1.



Pilt 1.

grupeeriva tunnuse koodide määramise näide on esitatud pildil 2. 1 ja 2 lahtrites määravad antud andmestikus meeste ja naiste koodid.



Pilt 2.

1) Kas keskmised hinnangud rahulolule oma senise eluga ja rahulolule oma materiaalse olukorraga on statistiliselt olulisel määral erinevad meestel ja naistel

Analyze – Compare Means – Independent-Samples T Test

☐ Grouping Variable: rahulolu tunnused

Define Groups: Sugu

2) Kas keskmine hinnang küsimusele „Kuivõrd mures Te olete, kui üldse, sellepärast, et Teie sissetulek vanemas eas ei ole piisav, et tagada Teie heaolu hilisemas elus“ sõltub sellest, kas

vastaja on juba vanaduspäeviks säästnud (Küsimus „Kas Te säästate praegu või olete varasemalt säästnud selleks, et vanemas eas mugavalt ära elada “)

Tulemuste esitamine

Keskmine hinnang valitsuse tegevusele ei erinenud soo lõikes, $t = 0.2$, $p = 0.240$.

($t = 0.2$, $p = 0.24$)

Naiste keskmine hinnang valitsuse tegevusele oli kõrgem kui meestel ($t = 5.65$, $p = 0.05$).
duses peab selguma indiviidide arv.

Tabel 1. Meeste ja naiste keskmised usaldushinnangud

	Keskmine	Standardhälve	N
Mehed	5.2	1.6	160
Naised	5.6	1.1	238
Kokku	5.4	1.7	398

Tabel 2. Meeste ja naiste keskmised usaldushinnangud

	Keskmine	Standardhälve	N
Mehed	5.2	1.6	160
Naised	5.6	1.1	238
t	1.2		
p	0.35		

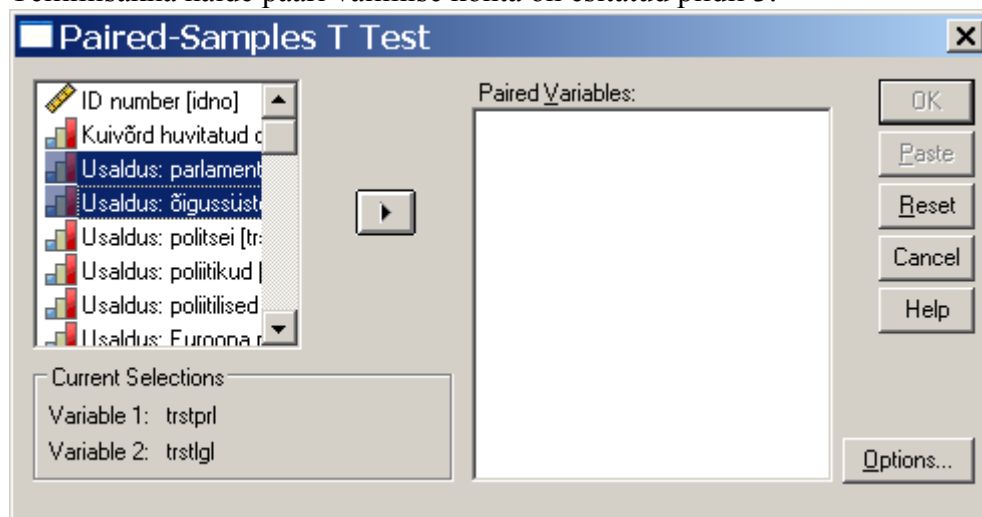
Kui tulemused on esitatud tabelina, siis peaks tekstis olema ka kindlasti viide tabelile.

Paariviisiline T-test

Analyze – Compare Means – Paired-Samples T Test

Paired Variables: valige tunnuspaar

Tellimisakna näide paari valimise kohta on esitatud pildil 3.



Pilt 3.

3) Selgitage välja, kumba usaldavad vastajad rohkem, kas parlamenti või õigussüsteemi.

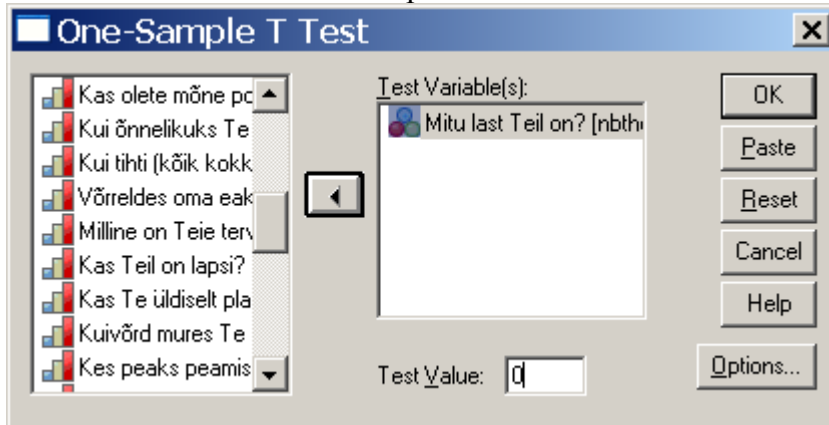
Analyze – Compare Means – Paired-Samples T Test

One Sample T-test

Analyze – Compare Means – One-Sample T Test

Kas testitav tunnus on erinev mingist konkreetsest väärtusest, näiteks 0 (vaikimisi valikuna tellimisaknas, kuid muudetav)

Tellimisakna näide on esitatud pildil 4.



Pilt 4.

4) Leidke, kas keskmine laste arv vastajatel on statistiliselt olulisel määral erinev 2-st

Analyze – Compare Means – One-Sample T Test

oonana

1. variant: üks sõltuv tunnus ja üks grupeeriv tunnus (faktor)

Analyze – Compare Means – One-Way ANOVA

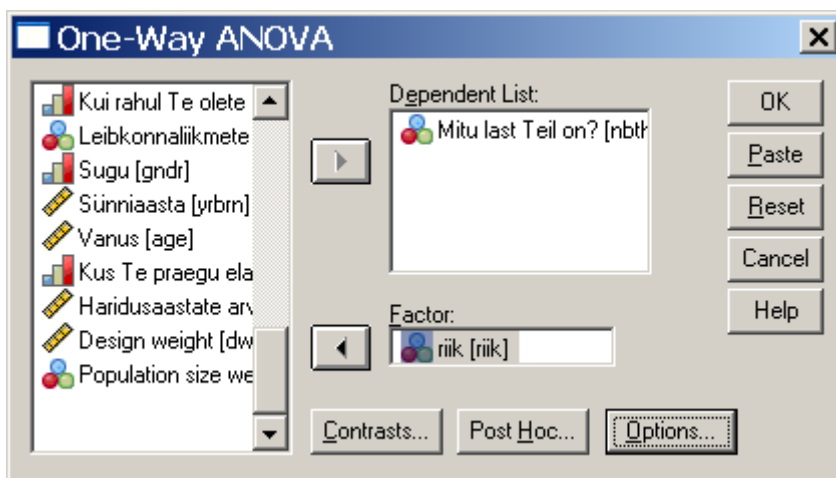
Dependent List: sõltuv tunnus (see, mille keskmist me tahame saada)

Factor: grupeeriv tunnus (faktor)

Post Hoc: paarikaupa võrdlused. Kui grupeerival tunnusel on rohkem kui 2 gruppi, siis võib neid gruppe omavahel paarikaupa võrrelda. Valida näiteks Scheffe.

Options: valida „Descriptives“ ehk kirjeldav statistika ja „Means plot“ ehk keskmiste joonis.

Näide sellise dispersioonanalüüsi tellimisakna kohta on antud pildil 5.



Pilt 5

5) Selgitage välja, kas vastajate laste arv on eri riikides statistiliselt olulisel määral erinev
☐ illises riigis on lastega peredes kõige rohkem lapsi

Analyze – Compare Means – One-Way ANOVA

Dependent List: ☐ itu last Teil on ☐

☐actor: riik

6) Kas keskmised hinnangud enda õnnelikkusele olid riigiti erinevad

Kuidas esitada tulemusi?

Keskmised hinnangud enda õnnelikkusele olid riigiti erinevad ($\chi^2(103) = p < 0.000$).

Keskmised hinnangud enda õnnelikkusele olid riigiti statistiliselt olulisel määral erinevad ($\chi^2(103) = p < 0.05$).

oonana

2. variant: üks sõltuv tunnus ja mitu grupeerivat tunnust

Analyze – General Linear Models – Univariate

Dependent Variable: sõltuv tunnus

☐ixed ☐actor(s): grupeerivad tunnused

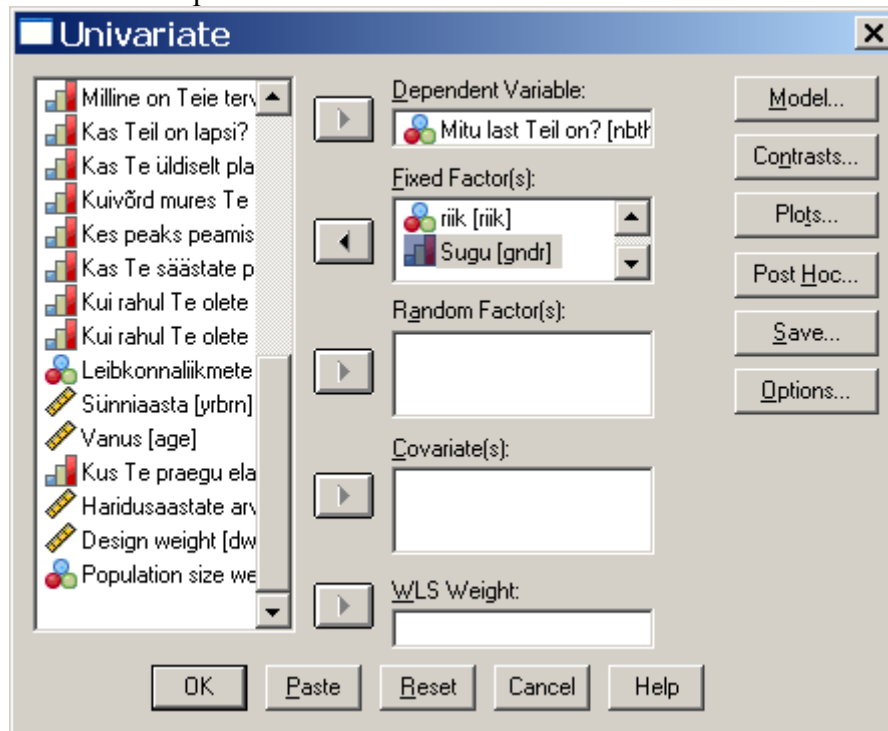
☐odel: saab valida täpsemalt mudeli ja koosmõjud

Plots: keskmiste joonised (nii üksiku kui koosmõju kohta). Iga joonis tuleb eraldi tellida, valides telgedel ja joontel näidatavad tunnused ning vajutades „Add“ nuppu.

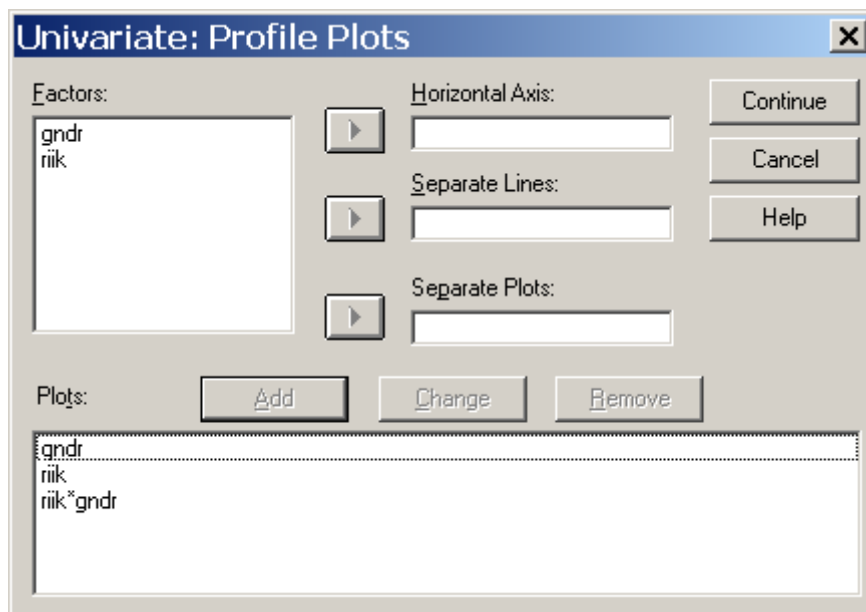
Post hoc: paarikaupa võrdlus (valida rohkem kui 2 grupiga tunnused ning testiks näiteks Scheffe)

Options: valida „Descriptive statistics“

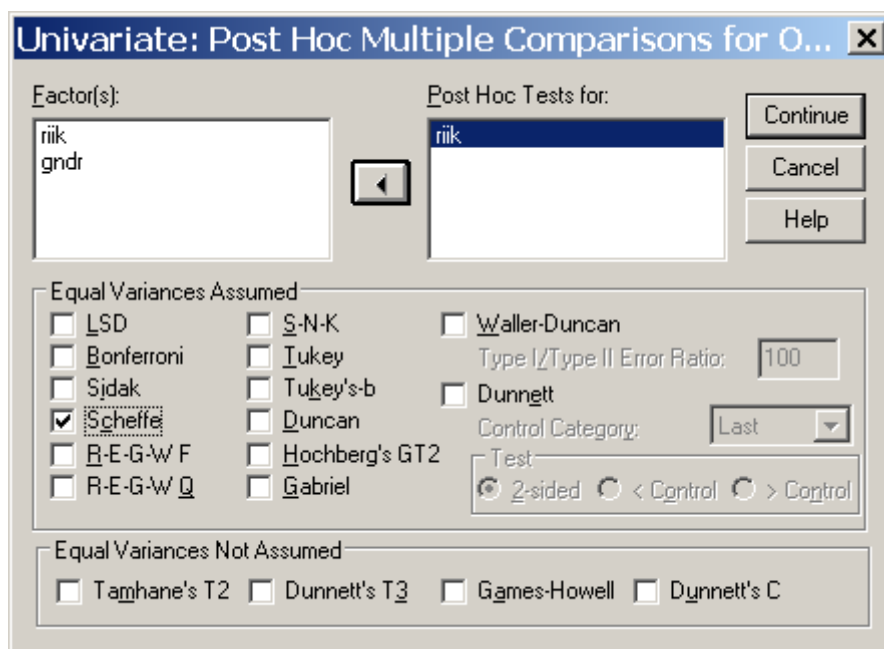
Peamine tellimisaken on näidatud pildil 6, keskmiste jooniste tellimine pildil 7 ning post hoc testi tellimine pildil 8.



Pilt 6.



Pilt 7.



Pilt 8.

7) Koostage enda valikul üks mitmefaktoriline dispersioonanalüüsi mudel, tehke joonised (ka koosmõjude kohta).

Statistikapaketi SPSS kasutamine

Praktikum □

Avage andmestik risttabel.sav.
Tutvuge tunnuste kirjeldustega

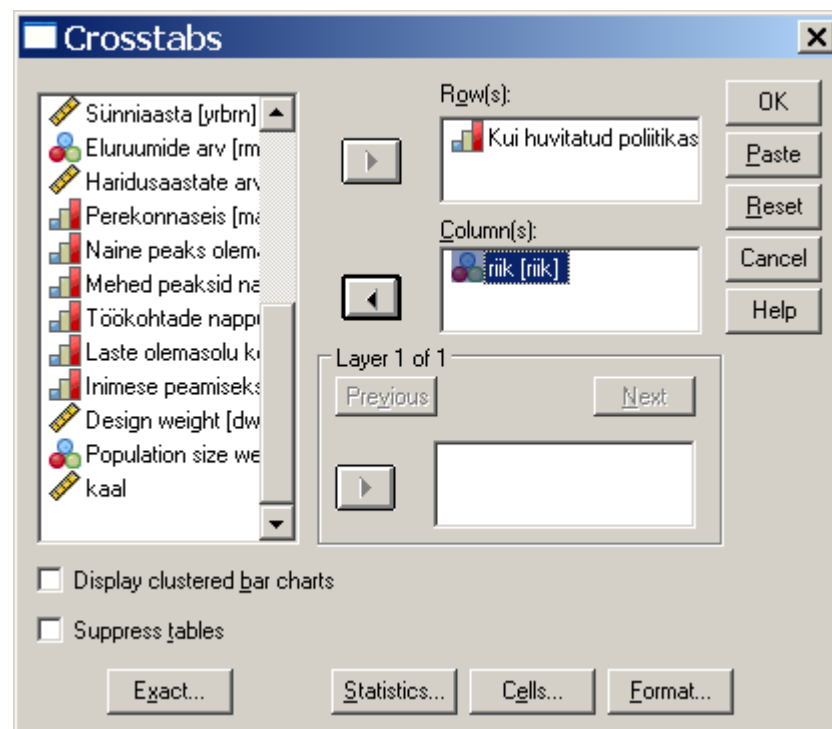
Risttabelid ehk mitmemõõtmelised sagedustabelid
Analyze – Descriptive Statistics - Crosstabs

Row(s): reatunnus
Column(s): veerutunnus
Layer: kihitunnus (eraldi tabelid)
Cells: protsendid ja inimeste arvud
Statistics: seosekordajad
 ormat: tabeli vormistus (kategooriate järgnevus ridades)

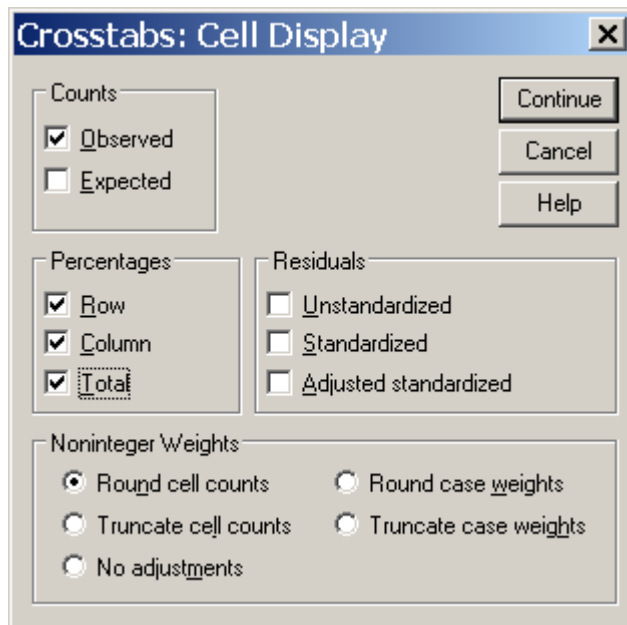
Reaprotsent (row): mitu protsenti selle rea inimestest kuulub ühte või teise veergu.
Veeruprotsent (column): mitu protsenti selle veeru inimestest kuulub ühte või teise ritta.

Idprotsendid (total): mitu protsenti selle tabeli inimestest kuulub ühte või teise lahtrisse.

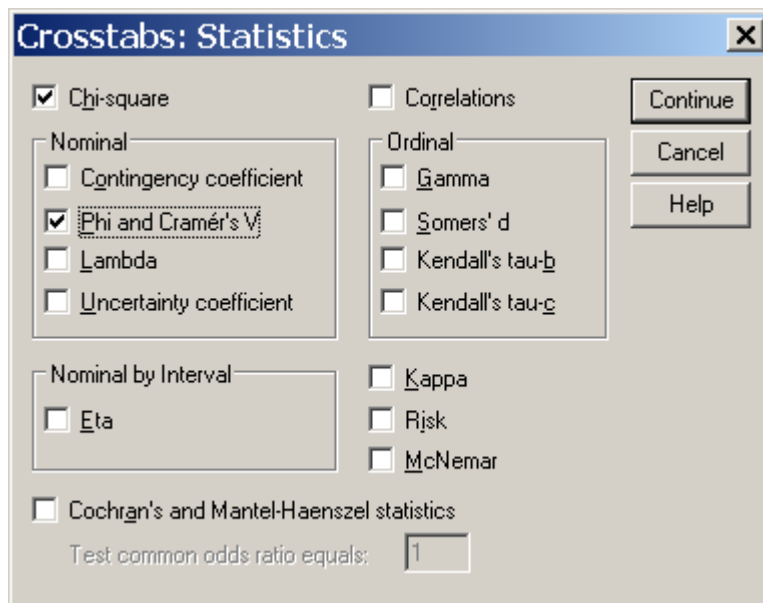
Risttabeli tellimisaken on esitatud pildil 1, protsentide valiku aken „Cells“ pildil 2 (valitud on kõik 3 võimalikku protsenti) ning seosekordajate valiku aken „Statistics“ pildil 3 (valitud hii□ruut□statistik, Phi ja Crameri V).



Pilt 1.



Pilt 2.



Pilt 3.

1) Tehke sagedustabel selle kohta, kuivõrd huvitatud on erinevate riikide vastajad poliitikast.

Analyze – Descriptive Statistics - Crosstabs

kavad edaspidi olema

saada risttabelid meeste ja naistele eraldi, siis saadame sinna aknasse soo tunnuse). Praegu ei pane sinna midagi.

2) Lisage eelmises ülesandes tellitud sagedustabelisse ka veeruprotsendid

Counts –
ning seda me ei muuda)
Percentages –

Percentages –

le

Percentages – Total annab protsendid terve tabeli kohta kokku (st mitu protsenti

- ☐ itu protsenti eestlastest on poliitikast väga huvitatud
illise riigi vastajad on poliitikast kõige rohkem huvitatud

3) Tellige sama tabel uuesti, kuid valige kõik võimalikud protsendid (Row, Column, Total)

- itu protsenti soomlastest ei tunne poliitika vastu üldse huvi
itu protsenti nendest inimestest, kes on poliitikast vähe huvitatud, elavad estis
itu protsenti poolakatest on poliitikast üsna huvitatud
☐ itu protsenti vastajatest elab ☐ estis ☐
illisest riigist on kõige rohkem vastajaid
itu protsenti kõigist vastajatest ei tunne poliitika vastu üldse huvi

4) Tehke risttabel riigi ja tunnuse „Kui tihti tundub poliitika nii keerulisena, et Te ei saa päriselt aru, mis toimub“ vahel. Püüdke sõnastada tulemusi ja järeldusi.

5) Tehke risttabel riigi ja tunnuse „Kas hääletas viimastel riiklikel valimistel“ vahel.
illise riigi vastajad käisid kõige rohkem ja millise riigi valijad kõige vähem valimas ☐

6) Tehke 3 ☐ mõõtmeline tabel selle kohta, kas poliitikahuvi ja valimas käimise tunnuste vaheline seos on erinevates riikides erinev.

Row(s): Kui huvitatud poliitikast

Column(s): Kas hääletas viimastel riiklikel valimistel

Layer: Riik

Püüdke sõnastada järeldusi

illises riigis sõltub poliitikahuvi kõige rohkem valimas käimisest (s.t millises riigis on kõige suuremad erinevused näiteks nende inimeste osakaaludes, kes on poliitikast väga huvitatud kui võrrelda valimaskäinuid ja valimas mittekäinuid.

7) Tehke sarnane tabel poliitikast arusaamise tunnuse kohta (Kui tihti tundub poliitika nii keerulisena, et Te ei saa päriselt aru, mis toimub). Kas seose muster on sarnane kui eelmise ülesande puhul

8) illises riigis tunnevad mehed kõige rohkem huvi poliitika vastu ja millises riigis tunnevad naised kõige rohkem huvi poliitika vastu Põhjendage vastust kasutades protsente.

- ☐ illine on olukord estis võrreldes teiste riikidega

) Koostage enda valikul üks 4-mõõtmeline tabel. (4. tunnuse jaoks Layer alt valida „Next“)

Seosekordajad

Analyze – Descriptive Statistics - Crosstabs

Nupu *Statistics* alt valida seosekordajad

Chi-square – hiiuutstatistik

Phi and Cramer's V

10) Selgitage välja, kas on erinevusi meeste ja naiste vastustes järgmistele küsimustele:

	Chi-square statistiku väärtus	Olulisuse tõenäosus	Crameri V	Olulisuse tõenäosus	Kas on statistiliselt oluline erinevus
Naine peaks olema valmis vähendama oma palgatööd oma pereelu nimel	16,7	0,002			
Mehed peaksid naistega võrdselt vastutama kodu ja laste eest			0,178	0,000	jah
Töökohtade nappuse korral peaks meestel olema eelisõigus tööd saada					
Laste olemasolu korral peaksid vanemad kokku jääma ka siis, kui nad ei saa hästi läbi					
Inimese peamiseks prioriteediks peaks olema tema perekond					

illise tunnuse puhul on seos sooga kõige tugevam

11) illise riigi vastajad on kõige rohkem nõus sellega, et inimese peamiseks prioriteediks peaks olema tema perekond Põhjendage oma vastust kasutades protsente. Kas seos riigi ja tunnuse „Inimese peamiseks prioriteediks peaks olema tema perekond“ vahel on statistiliselt oluline

12) illise riigi vastajad on kõige vähem nõus sellega, et mehed peaksid naistega võrdselt vastutama kodu ja laste eest Põhjendage oma vastust kasutades protsente.

13) Kas inimeste hinnangud küsimusele „Inimese peamiseks prioriteediks peaks olema tema perekond“ sõltuvad nende perekonnaseisust Kui jah, siis kuidas? Kas see seos on sarnane nii meestel kui naistel?

Statistikapaketi SPSS kasutamine

Praktikum 10. Korrelatsioonikordajad ja hajuvusdiagrammid

Avage andmestik maakonnad.sav

Tutvuge tunnuste kirjeldustega

Hajuvusdiagrammid

Graphs – Legacy Dialogs – Scatter/Dot

Simple Scatter

Axis: püstteljel kujutatav tunnus

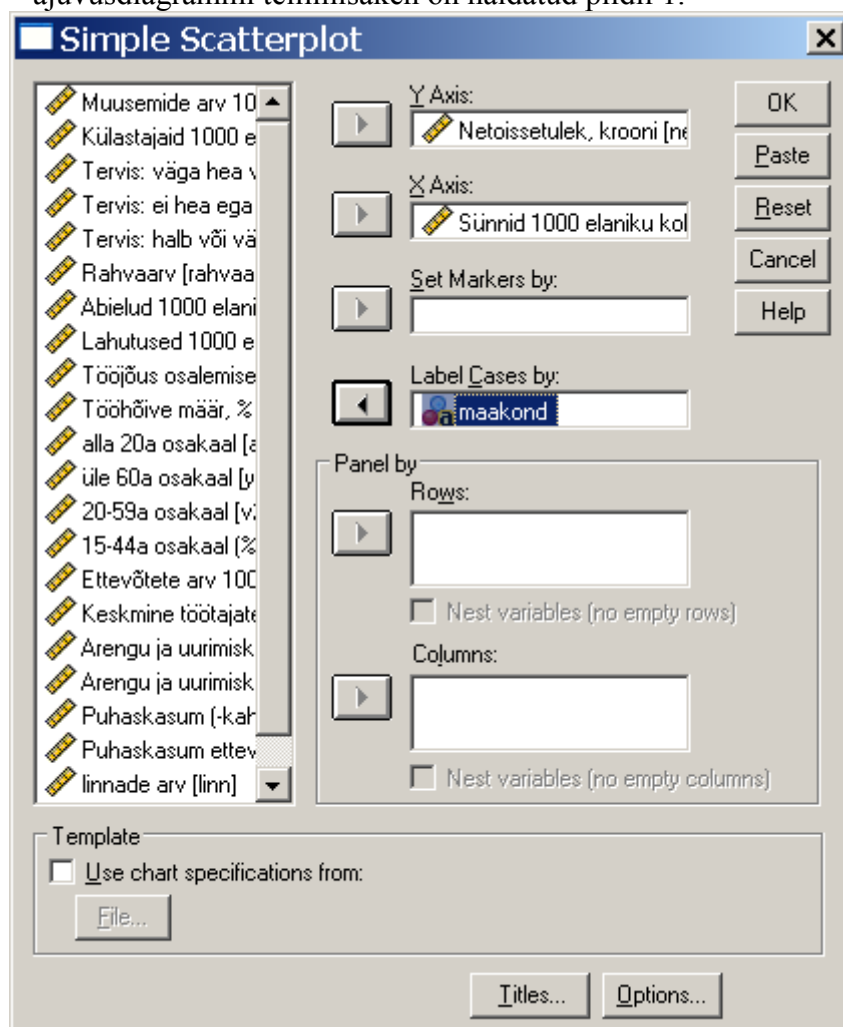
☐ Axis: ristteljel kujutatav tunnus

Set markers y: võimaldab märgistada joonisel olevad punktid vastavalt mingi kolmanda tunnuse väärtusele

Label Cases y: võimaldab anda joonisel olevatele punktidele nimed, kui on olemas vastav nimetunnus

Options: puuduvate väärtuste näitamine, kustutamine nimetunnuse olemasolul selle näitamine

ajuvusdiagrammi tellimisaken on näidatud pildil 1.



Pilt 1.

1) Tehke hajuvusdiagramm tunnuste „Netosissetulek“ ja „Sünnid 1000 elaniku kohta“ põhjal. Kas sündide arvu ja maakonna elanike netosissetuleku vahel on seos? Kui jah, siis milline?
Graphs – Legacy Dialogs – Scatter/Dot

Joonise redigeerimisel saab sinna lisada ka trendijooone (regressioonijooone) ning determinatsioonikordaja. Selleks tuleks teha topeltklõps joonisel (redigeerimisakna avamiseks) ning valida *Elements – Fit Line at Total* või valida nupp „Fit line at total“ Sarnase tulemuse saate, kui valite nupu „Add a reference line from equation“, selle variandiga saate ka ise joone valemi ette anda.

2) Tehke hajuvusdiagramm netosissetuleku ja sündimuse kohta, kuid lisage ka markerina linnade arv maakonnas.

3) Tehke hajuvusdiagramm netosissetuleku ja sündimuse kohta, lisage ka punktidele maakondade nimed.

4) Uurige hajuvusdiagrammi abiga, kas on olemas seosed netosissetuleku ja abiellumise ning netosissetuleku ja lahutuste arvu vahel.
Joonisel paistab silma üks erandlik maakond, mille poolest see ülejäänutest eristub

5) Tehke 3 mõõtmeline hajuvusdiagramm netosissetuleku, sündide ja 15–44a inimeste osakaalu kohta maakondades.

6) Tehke omal valikul veel vähemalt kolm hajuvusdiagrammi andmestikus olevate tunnuste põhjal ning püüdke sõnastada järeldusi.

Korrelatsioonikordaja

Pearsoni ehk lineaarne korrelatsioonikordaja (arvtunnuste vahel)

Spearmani astakorrelatsioonikordaja (arv- või pikema skaalaga järjestustunnuste vahel)

Vahemikus -1 kuni 1

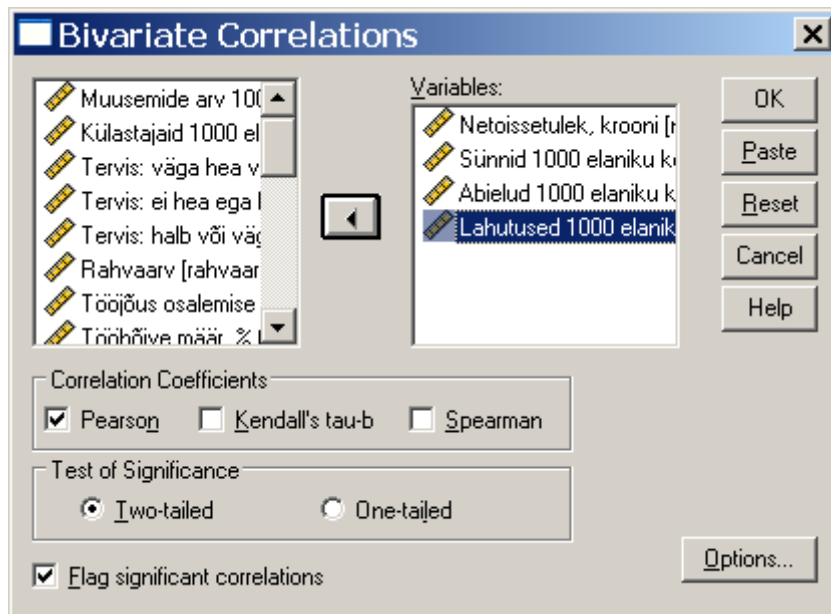
aksimaalne seosetugevus absoluutväärtusega 1, minimaalne 0

Analyze – Correlate – Bivariate

Variables: korrelatsioonimaatriksisse soovitatavate tunnuste nimekiri (vähemalt 2)

Tunnuste akna all on võimalik valida korrelatsioonikordajat (vaikimisi Pearsoni ehk lineaarne korrelatsioonikordaja)

Tellimisaken on näidatud pildil 2.



Pilt 2.

7) Tehke korrelatsioonimaatriks 4 tunnuse: netoissetulek, sündide arv, abielude arv ja lahutuste arv – omavaheliste seoste kohta.

8) Kui andmestikust välja jätta arjuma, siis kuidas korrelatsioonikordajad muutuvad
Proovige järgi (kasutage filtrit Data □Select Cases)

) Kuidas muutuvad korrelatsioonikordajad, kui arjuma asemel jätta andmestikust välja Ida□Virumaa□

10) Leidke vähemalt 3 näitajat, mis on positiivselt korreleeritud ettevõtete arvuga 1000 elaniku kohta maakonnas, vähemalt 2 näitajat, mis on sellega negatiivselt korreleeritud ning 1 näitaja, mille puhul korrelatsioonikordaja on praktiliselt 0.

11) Kumb tunnus on netoissetulekuga tugevamini seotud, kas „üle 60a osakaal“ või „Lahutused 1000 elaniku kohta“

Avage andmestik esskvm.sav

12) Tehke korrelatsioonimaatriks järgmiste tunnuste vaheliste korrelatsioonseoste kohta, kasutage Spearmani korrelatsioonikordajat.

Tunnused:

- Kui õnnelikuks Te kõike kokkuvõttes end peate
- Kui tihti (kõik kokku) kohtute veedate aega sõprade, sugulaste või töökaaslastega
- Võrreldes oma eakaaslastega, kui tihti Te võtate osa seltskondlikest kohtumistest või üritustest
- Kui rahul Te olete oma praeguse materiaalse olukorraga□

□ illine nimetatud tunnustest on õnnelikkuse hinnanguga kõige tugevamalt seotud □ illine see seos on□

Kirjeldage ka teisi leitud seoseid. Kas seosed õnnelikkuse hinnanguga on statistiliselt olulised□ tugevad või nõrgad □ positiivsed või negatiivsed.

Vormistage tulemus tabelina ja korrelatsioonigraafina (sobiv seosetugevus valige ise). Näidised on esitatud järgnevalt

Tulemuste esitamine

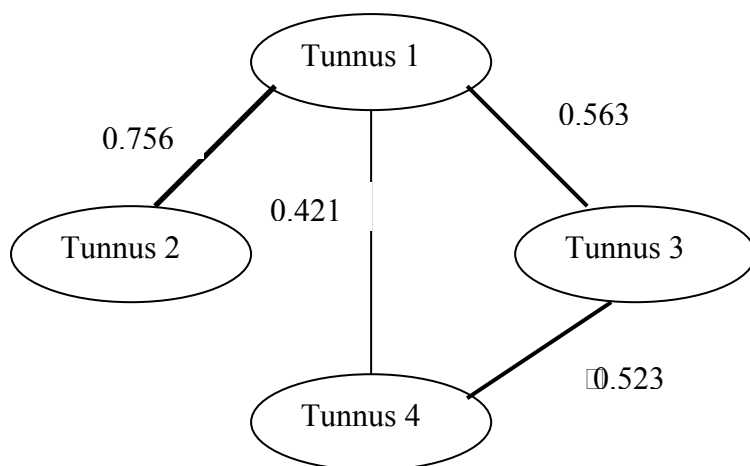
Korrelatsioonimaatriks

	Tunnus 1	Tunnus 2	Tunnus 3	Tunnus 4
Tunnus 1	1			
Tunnus 2	0.756 <input type="checkbox"/>	1		
Tunnus 3	0.563 <input type="checkbox"/>	0.102	1	
Tunnus 4	0.421 <input type="checkbox"/>	0.21 <input type="checkbox"/>	-0.523 <input type="checkbox"/>	1

☐ $p < 0.001$ ☐ $p < 0.01$ ☐ $p < 0.05$

Tabeli juures märkida ka ära, millise korrelatsioonikordajaga (lineaarne, Spearmani vm) on tegemist ja kui suur on vastajate arv.

Korrelatsioonigraaf (märgitud seosed, mis on vähemalt tugevusega 0.4)



13) Selgitage välja vähemalt 3 tegurit, mis on seotud vastajate hinnanguga oma tervislikule seisundile. Sõnastage järeldused.

☐ illist korrelatsioonikordajat tuleks kasutada ☐

14) Tehke eelmises ülesandes kasutatud korrelatsioonimaatriksid erinevate riikide jaoks. Selleks on vaja andmestik jagada (Split file).

Statistikapaketi SPSS kasutamine

Praktikum 11. S ntak ka ta n

Avage andmestik ess1.sav

Uue programmeerimisakna avamiseks

File – New - Syntax

Olemasoleva programmeerimisakna avamiseks

File – Open – Syntax

1) Avage fail syntax1.sps

Selles on näidatud mõned SPSS programmeerimiskäsud, uurige neid kõigepealt.

Pange tähele, et SPSS käsud lõppevad punktiga.

Käsu käivitamiseks võtta see hiirega blokki ja seejärel üks järgnevatest võimalustest:

- Vajutada sinise noolega nupule
- Parema hiireklahviga Run current
- enüü Run

ärkmed, mida soovite teha süntaksisse, kuid mida ei soovi käivitada, tuleks panna selliselt, et neid sisaldavad read algaksid tärniga.

Näiteks:

ärkmed käskude kohta

SPSSi väljundit saab seadistada selliselt, et seal (Output aknas) näidataks ka käivitatud käskusid.

Praktiliselt kõigis tellimisakendes on olemas nupp „Paste“, mis trükitab tellitava käsu süntaksiaknasse, kuid ei käivita seda veel. Proovige järgi.

2) Tehke sagedustabel tunnuse „Sugu“ kohta, seejuures ärge tellimisaknas (Analy e – Descriptive Statistics – re uencies) valige OK, vaid Paste.

namasti annab tellimisaknast võetud käske ka lühendada. Seda on tehtud näiteks faili syntax1.sps algustabelite puhul.

3) Vaadake järgi levinumate tabelite ja jooniste tellimise süntaksi käsud (tellides tabelid menüüdest ning vajutades „Paste“).

4) Kodeerige mõni andmestikus olemasolev tunnus ümber uueks tunnuseks. Vaadake süntaksit (Transform – Recode Into Different Variables)

5) oodustage andmestikus olemasolevatest tunnustest mõni uus tunnus. Vaadake süntaksit (Transform – Compute Variables).

6) Avage fail syntax vigadega.sps ja proovige sealt leida vigu süntaksi kirjutamises.

n d k d

Faili avamine

□□T □IL□

GET FILE='c:\data\empl.sav'.

Tunnustele nime panemine

VARIA□L□LA□□LS

VARIABLE LABELS YRHIRED 'YEAR OF FIRST HIRING'.

VARIABLE LABELS YRHIRED 'YEAR OF FIRST HIRING'
DEPT88 'DEPARTMENT OF EMPLOYMENT IN 1988'
SALARY88 'YEARLY SALARY IN 1988'
JOBCAT 'JOB CATEGORIES'.

VALU□LA□□LS

VALUE LABELS JOBGRADE 'P' 'Parttime Employee' 'C' 'Customer Support'.

VALUE LABELS V1 TO V3 1 'Officials & Managers'
6 'Service Workers'
/V4 'N' 'New Employee'.

□ISSIN□ VALU□S

MISSING VALUES V1 (8,9) V2 V3 (0) V4 ('X').
MISSING VALUES V1 (LOWEST THRU 0).
MISSING VALUES V1 (0 THRU 1.5).
MISSING VALUES V1 (LO THRU 0, 999).

Uue tunnuse moodustamine

CO□PUT□

COMPUTE V1=25-V2.
COMPUTE V3=(V2/V4)*100.

R□COD□

RECODE V1 TO V3 (0=1) (1=0) (2,3=-1) (9=9) (ELSE=SYSMIS).
RECODE STRNGVAR ('A','B','C'='A') ('D','E','F'='B') (ELSE=' ').

Kirjeldav statistika

□R□□U□NCI□S

FREQUENCIES VARIABLES = RACE.
FREQUENCIES VARIABLES = RACE /STATISTICS=ALL.

CROSSTA□S

CROSSTABS TABLES=FEAR BY SEX
/CELLS=ROW COLUMN EXPECTED RESIDUALS

```
/STATISTICS=CHISQ.
```

```
□ □ANS
```

```
MEANS TABLES=V1 TO V5 BY GROUP.
```